

Multiway Online Correlated Selection

Guy Blanc

Moses Charikar

Stanford University

Stanford University

September 22, 2021

Abstract

We give a 0.5368-competitive algorithm for edge-weighted online bipartite matching. Prior to our work, the best competitive ratio was 0.5086 due to Fahrback, Huang, Tao, and Zadimoghaddam (FOCS 2020). They achieved their breakthrough result by developing a subroutine called *online correlated selection* (OCS) which takes as input a sequence of pairs and selects one item from each pair. Importantly, the selections the OCS makes are negatively correlated.

We achieve our result by defining *multiway* OCSes which receive arbitrarily many elements at each step, rather than just two. In addition to better competitive ratios, our formulation allows for a simpler reduction from edge-weighted online bipartite matching to OCSes. While Fahrback et al. used a factor-revealing linear program to optimize the competitive ratio, our analysis directly connects the competitive ratio to the parameters of the multiway OCS. Finally, we show that the formulation of Fahrback et al. can achieve a competitive ratio of at most 0.5239, confirming that multiway OCSes are strictly more powerful.

1 Introduction

The matching problem has played a pivotal role in the development of algorithmic techniques in combinatorial optimization. The online version of the problem was one of the first graph optimization problems studied in the online literature. Karp, Vazirani and Vazirani [KVV90] initiated a study of the problem and gave a $1 - 1/e$ competitive algorithm for unweighted graphs. Here, the vertices of one side (the offline side) are known to the algorithm in advance. The vertices of the other side appear in an online fashion. As they appear, they must be matched irrevocably to an offline vertex (or discarded). Since the work of [KVV90], online bipartite matching received considerable attention and their ideas were applied to a variety of other online assignment problems.

More than a decade later, the emergence of sponsored search auctions gave renewed impetus to the study of online matching and its variants. There is a natural correspondence here – the offline side corresponds to advertisers and the online side corresponds to impressions (i.e. opportunities to place ads). Aggarwal et al [AGKM11] generalized the [KVV90] result to the vertex-weighted case, obtaining a $1 - 1/e$ competitive algorithm.

The edge-weighted case had remained a tantalizing open problem until very recently. Here, it is not hard to see that no non-trivial competitive ratio is possible without giving the algorithm additional flexibility. Researchers have focused on the *free disposal* setting where offline vertices can be assigned multiple online vertices but the contribution of an offline vertex to the objective function is the largest of the edge weights assigned to that vertex. (i.e. previously matched edges can be disposed for free to make room for higher weight edges). In the display advertising setting, this corresponds to assigning multiple impressions to an advertiser.

For the edge weighted problem, a natural greedy algorithm achieves competitive ratio 0.5. This was the best known until the recent breakthrough by Fahrbach, Huang, Tao and Zadimoghaddam [FHTZ20] who gave a 0.5086 competitive algorithm. One of their main technical contributions was a novel algorithmic ingredient called *online correlated selection* (OCS). This is an online subroutine that takes a sequence of pairs of vertices as input and selects one vertex from each pair such that decisions across pairs are negatively correlated. A suitable quantification of this negative correlation was plugged into a factor revealing linear program to obtain a bound on the competitive ratio.

In this work we generalize the OCS definition of Fahrbach et al [FHTZ20] to multiway OCSes which take in multiple vertices at each step. Our analysis directly connects the parameters of our multiway OCS to the eventual competitive ratio without the need to go through a factor revealing linear program. We give a general framework for obtaining multiway OCSes and instantiate this framework to obtain an improved competitive ratio of 0.5368 for online weighted matching. Finally, we show that the OCS definition of Fahrbach et al [FHTZ20] cannot lead to a competitive ratio better than 0.5239 demonstrating that our multiway OCSes are strictly more powerful.

1.1 Related Work

Online weighted matching has been extensively studied in the literature. See the excellent survey by Mehta [Meh13] for a comprehensive overview. Two popular settings where competitive ratios better than 0.5 have been obtained are

- (1) the case where offline vertices have large capacity (i.e a large number of online vertices can be assigned to an offline vertex). This includes work of Kalyanasundaram and Pruhs [KP00], Feldman et al [FKM⁺09] on Display Ads and work on the Adwords problem [MSVV07, BJN07].
- (2) the case where stochastic information is available about the online vertices. This includes work

in the setting where online vertices are drawn from a known or unknown distribution [FMMM09, KMT11, DJSW19, HMZ11, MGS12, MP12, JL14] or the setting that they arrive in a random order [GM08, DH09, FHK⁺10, MY11, MGZ12, MWZ14, HTWZ19].

In addition to these, several recent advances have been made in more general settings including non-bipartite graphs and different arrival models [HKT⁺18, GKS19, GKM⁺19, HPT⁺19].

1.2 Problem definition: Edge-weighted online bipartite matching

Consider some weighted bipartite graph $G = (L, R, w)$, where L and R are the left and right vertices respectively. If there is an edge between $i \in L$ and $j \in R$, then $w_{ij} > 0$ is the weight of that edge. Otherwise, $w_{ij} = 0$.

At the start, the algorithm is given the entire set of left vertices, L , but no information about R or w . The vertices from R appear in an online fashion, one by one. Hence, we refer to L as the *offline* vertices R as the *online* vertices. When an online vertex $j \in R$ appears, the entries w_{ij} for each $i \in L$ are revealed to the algorithm. The algorithm must irrevocably decide which offline vertex to match j to before the next online vertex appears.

The objective is to maximize the total weight of the matching. We operate in the *free disposal model*. This means a single offline vertex i may be matched to multiple online vertices, but only the weight of its heaviest edge is counted towards the objective. We say a randomized algorithm is “ Γ -competitive” or “has competitive ratio of Γ ” if the expected objective of the algorithm’s output is within a multiplicative factor of Γ of the optimal objective with hindsight.

2 Online correlated selection: Motivation and application

We withhold a full preliminaries section until [Section 6](#), mentioning for now that we use **boldface** (e.g. $\mathbf{x} \sim \mathcal{D}$) to denote random variables.

It is well-known that even for *unweighted* online bipartite matching, deterministic algorithms cannot achieve a competitive ratio better than 0.5. In this section, we’ll investigate what form of randomness is needed to push the competitive ratio beyond 0.5.

At the j^{th} time step, the portion of the graph corresponding to $j \in R$ is revealed to the algorithm and it must irrevocably match it to some $\mathbf{i}_j \in L$. In order to achieve a competitive ratio better than 0.5, the choices $\mathbf{i}_1, \mathbf{i}_2, \dots$ must use randomness. What if those choices are independent?

Lemma 2.1. *Even for unweighted online bipartite matching, no algorithm making decisions that are independent across time steps can achieve a competitive ratio better than 0.5.*

We include a proof of [Lemma 2.1](#) in [Appendix A](#). In order to break the 0.5-competitive barrier, not only does the algorithm need to carefully select a marginal distribution for each \mathbf{i}_j , it also needs some “smarts” in deciding how to realize \mathbf{i}_j — they cannot simply be independent. The key insight underlying Online Correlated Selection (OCS) is that these tasks can be tackled individually.

We separate the online algorithm into two components. The first component is fully deterministic and has access to the input instance. At the j^{th} time step, it chooses a marginal distribution for \mathbf{i}_j as a deterministic function of the portion of the graph revealed so far. In particular, that marginal distribution does not depend on the realizations of $\mathbf{i}_1, \dots, \mathbf{i}_{j-1}$.

The second component, the OCS, is given the marginal distribution \mathbf{i}_j , but *not* given any other information about the input instance — It cannot “directly see” the graph. At the j^{th} time step, it

At time step $j \dots$	Deterministic Component	Randomized Component (OCS)
Information revealed	The edge weights, w_{ij} for each $i \in L$	The marginal distribution for \mathbf{i}_j (from the deterministic component) and the values for \mathbf{i}'_j for each $j' < j$
Information hidden	The randomized selections of the OCS	The input instance
Output	A marginal distribution for \mathbf{i}_j	A realization for \mathbf{i}_j

Figure 1: We separate a matching algorithm into a deterministic component and randomized component, each with different responsibilities and access to information.

chooses a realization for \mathbf{i}_j consistent with the desired marginal as a function of its own randomness. In particular, the realization for \mathbf{i}_j can depend on the previous realizations of $\mathbf{i}_1, \dots, \mathbf{i}_{j-1}$. This separation of responsibilities is summarized in [Figure 1](#).

In the following definition, a *probability vector* is a nonnegative vector whose elements sum to 1. It represents a marginal distribution over elements in L .

Definition 1 (Continuous OCS). *Consider a set of ground elements, L . A continuous OCS is an online algorithm that at each time step $j = 1, 2, \dots$, receives a probability vector $p^{(j)}$ with $|L|$ elements and irrevocably chooses a winner in $\mathbf{i}_j \in L$ among those with $(p^{(j)})_i > 0$.*

By [Lemma 2.1](#), in order to achieve a competitive ratio of greater than 0.5, the OCS cannot simply choose winners independently at each round according to the desired marginal distribution. In the next subsection, we examine what properties an OCS needs to have to allow for better competitive ratios.

2.1 What makes a good OCS?

The algorithm’s goal is to maximize the expected weight of the matching over the randomness of the OCS. Having many online vertices $j \in R$ matched to the same offline vertex $i \in L$ can be wasteful. This is because only the heaviest match counts for the objective. To counteract this, the choices the OCS makes will be *negatively correlated*. In particular we wish for, $j \neq j' \in R$, that:

$$\Pr[\mathbf{i}_j = \mathbf{i}_{j'} = i] < (p^{(j)})_i \cdot (p^{(j')})_i$$

Unfortunately, it is impossible for the OCS to negatively correlated decisions across all time steps by any non-negligible amount. We formalize that impossibility in [Lemma A.1](#), proved in [Appendix A](#). To get around that, we negatively correlated time steps that are “close” temporally. Let $S = \{j_1, j_1 + 1, \dots, j_2\}$ be a set of *consecutive* time steps. We’ll require our OCS to satisfy

$$\Pr[\mathbf{i}_j \neq i \text{ for every } j \in S] \leq f \left(\sum_{j \in S} p_i^{(j)} \right)$$

where $f : \mathbb{R}_{\geq 0} \rightarrow [0, 1]$ is a function quantifying the negative correlation of an OCS. The OCS that chooses winners independently at each round according to the marginal distribution satisfies

the above equation with $f_{\text{trivial}}(x) := e^{-x}$. OCSes with “good” negative correlation have $f(x) < f_{\text{trivial}}(x)$ and the amount f is smaller than f_{trivial} directly connects to competitive ratio of the resulting algorithm.

The selections at time steps that are very far apart temporally will essentially be independent, as in the following definition.

Definition 2 (Quantifying a continuous OCS, informal version of [Definition 12](#)). *For any function $f : \mathbb{R}_{\geq 0} \rightarrow [0, 1]$, an f -continuous OCS is a continuous OCS with the following guarantee. For any element $i \in L$ and S_1, \dots, S_k each sets of consecutive time steps,*

$$\Pr[\mathbf{i}_j \neq i \text{ for every } j \in S_1 \cup \dots \cup S_k] \leq \prod_{\ell=1}^k f\left(\sum_{j \in S_k} (p^{(j)})_i\right).$$

We give an algorithm for edge-weighted online bipartite matching using an f -OCS as a subroutine and are able to directly connect the competitive ratio to f .

Theorem 1 (Informal version of [Theorem 6](#)). *For any differentiable and convex $f : \mathbb{R}_{\geq 0} \rightarrow [0, 1]$, if there is an f -continuous OCS, there is an algorithm for edge-weighted online bipartite matching with competitive ratio of*

$$\Gamma := 1 - \int_0^\infty e^{-t} f(t) dt.$$

As expected, using $f_{\text{trivial}}(x) := e^{-x}$ gives a competitive ratio of 0.5. Any f -OCS for f strictly less than f_{trivial} gives a competitive ratio that breaks the 0.5 barrier.

The proof of [Theorem 1](#) uses a similar online primal-dual formulation as that of [\[FHTZ20\]](#). We find that using our definition for an OCS in place of theirs substantially simplifies that analysis. In particular, we derive a simple analytic formula for the competitive ratio whereas they use a factor-revealing LP. The source of that simplification is that their algorithm needs to decide which of three round types, “deterministic,” “randomized,” or “unmatched,” to execute at each time step. Our continuous formulation implicitly interpolates between all three round types.

2.2 Discrete OCSes

We define discrete OCSes for two reasons.

1. It allows us to relate our results to those of [\[FHTZ20\]](#) who used (what we call) a discrete OCS.
2. Later, to construct a continuous OCS, we will first construct a discrete OCS and use a reduction to turn it into a continuous OCS.

The following definition of an m -discrete OCS is equivalent to a continuous OCS where each element in the probability vector is forced to be an integer multiple of $\frac{1}{m}$. Therefore, continuous OCSes are generalizations of discrete OCSes.

Definition 3 (Discrete OCS). *Consider a set of ground elements, L . An m -discrete OCS is an online algorithm that at each time step $j = 1, 2, \dots$, is given a size- m multiset A_j of elements in L and irrevocably selects a single winner $\mathbf{i}_j \in A_j$.*

We quantify the quality of a discrete OCS similarly as for continuous OCSes.

Definition 4 (Quantifying a discrete OCS). *For any $m \in \mathbb{Z}_{\geq 2}$ and $F : \mathbb{Z}_{\geq 0} \rightarrow [0, 1]$, an (F, m) -discrete OCS is an m -discrete OCS with the following guarantee. For any element $i \in L$ and S_1, \dots, S_k each sets of consecutive time steps,*

$$\Pr[i_j \neq i \text{ for every } j \in S_1 \cup \dots \cup S_k] \leq \prod_{\ell=1}^k F \left(\sum_{j \in S_k} \text{count}_i(A_j) \right).$$

where $\text{count}_i(A_j)$ is the number of times i appears in the multiset A_j .

For example, the OCS that independently and uniformly i_j from A_j at each round achieves $F_{\text{trivial}}(n) = (1 - 1/m)^n$. Our goal is to construct OCSes with $F < F_{\text{trivial}}$, which directly leads to competitive ratios better than $\frac{1}{2}$.

Theorem 2 (Informal version of [Theorem 7](#)). *For any $m \in \mathbb{N}$ and convex $F : \mathbb{Z}_{\geq 0} \rightarrow [0, 1]$, if is there an (F, m) -discrete OCS, there is an algorithm for edge-weighted online bipartite matching with competitive ratio of*

$$\Gamma := 1 - \sum_{n=0}^{\infty} \frac{m^n}{(m+1)^{n+1}} \cdot F(n).$$

We prove (a formal version of) [Theorem 2](#) by showing how to construct a continuous OCS from a discrete OCS and then applying [Theorem 1](#). This is done in [Section 8](#).

Lastly, we connect our OCS definition to that of Fahrback et al.

Definition 5 (γ -OCS, Definition 2 of [\[FHTZ20\]](#)). *For any $\gamma \in (0, 1)$, a γ -OCS is an $(F, 2)$ -discrete OCS where*

$$F(n) := 2^{-n} \cdot (1 - \gamma)^{\max(n-1, 0)}.$$

As a straightforward application of [Theorem 2](#), we derive an explicit formula mapping γ to a competitive ratio.

Theorem 3. *For any $\gamma \in (0, 1)$, if there is a γ -OCS, there is an algorithm for edge-weighted online bipartite matching with a competitive ratio of*

$$\Gamma = \frac{3 + 2\gamma}{6 + 3\gamma}. \tag{1}$$

We experimentally compared the competitive ratio from [Equation \(1\)](#) to those of the gain-sharing LP used in [\[FHTZ20\]](#). For every $\gamma \in \{0, 0.01, \dots, 0.46\}$, the competitive ratios are within an additive 10^{-8} of one-another. That small difference is likely due to rounding issues. For $\gamma \in \{0.47, 0.48, \dots, 1\}$, our equation gives a greater competitive ratio. Note that, as we later prove in [Lemma 2.2](#), γ -OCSes can only exist for $\gamma \leq \frac{1}{3}$ and our competitive ratio appears to match theirs on that range.

Lastly, we remark that “perfect negative correlation” for [Definition 5](#) corresponds to $\gamma = 1$. As noted in [\[FHTZ20\]](#) and can be confirmed by plugging $\gamma = 1$ into [Theorem 3](#), that corresponds to a competitive ratio of $\Gamma = \frac{5}{9}$. This is far from the best known upper bound of $\Gamma = 1 - \frac{1}{e}$. On the

other hand, perfect negative correlation for a continuous OCS corresponds to $f(x) = (1-x)_+$ which gives a competitive ratio of $\Gamma = 1 - \frac{1}{e}$. While perfect negative correlation is impossible, the fact that it corresponds to a competitive ratio equal to the best known upper bound seems to suggest that continuous OCSes are a step in the right direction towards finding the optimal competitive ratio.

2.3 Negative results

We prove the following for Fahrbach et al.’s formulation of an OCS.

Lemma 2.2. *No γ -OCS exists for $\gamma > \frac{1}{3}$.*

Both [Theorem 3](#) and the factor-revealing LP of [\[FHTZ20\]](#) agree that the competitive ratio for $\gamma = \frac{1}{3}$ is less than 0.5239. This is an upper bound on the competitive ratios that can be achieved simply by constructing better γ -OCSes. Our competitive ratio of 0.5368 is larger, confirming that multiway OCSes are strictly more powerful.

3 Constructing a multiway OCS: Meta algorithm overview

We give a meta algorithm for constructing m -discrete OCSes given a “win distribution,” \mathcal{W} , over $[0, 1]^N$. A sample from \mathcal{W} is a sequence of probabilities, $(\mathbf{x}_1, \mathbf{x}_2, \dots)$, with the quality of the resulting OCS (i.e. the function F in [Definition 4](#)) depending on how negatively correlated the elements in that sequence are. During initialization, the meta algorithm draws an independent “win sequence”, $(\mathbf{x}_1(i), \mathbf{x}_2(i), \dots)$ for each $i \in L$. Our OCS also stores an index within the win distribution, $k(i)$ for each $i \in L$ that is initialized to $k_i = 1$.

At each time step, the OCS receives $A_j = \{i_1, \dots, i_m\}$. For each $i \in A_j$, it determines the “desired win probability” of i in this time step. In the case where i appears exactly once in A_j , this desired win probability is exactly equal to $\mathbf{x}_{k(i)}(i)$. When i appears $r \geq 2$ times in A_j , the desired win probability is the probability i would win in at least one of the next r rounds if it appeared exactly once in each. Then, a tournament algorithm takes in these desired win probabilities and selects a single winner for the round consistent with the desired win probabilities and $k(i)$ is incremented r times. We include pseudocode for our meta algorithm in [Figure 2](#), though defer description of the win distribution and tournament subroutine for later.

The fact that the desired win probabilities of different vertices are *independent* makes analysis tractable. To achieve a competitive ratio better than 0.5, we need the winners of our OCS to be negatively correlated. We achieve this by making the random variables in the sequence returned by the win distribution negatively correlated. This way, if i has a small chance of winning in one round, it has a larger chance of winning in other rounds. In particular, we need two types of negative correlation for $(\mathbf{x}_1, \mathbf{x}_2, \dots) \sim \mathcal{W}$.

1. A *qualitative* and *long-distanced* negative correlation. We need to ensure there doesn’t exist any positive correlation between \mathbf{x}_{k_1} and \mathbf{x}_{k_2} even for k_1 and k_2 far apart. We prove that our win probabilities satisfy a generalization of negatively correlated random variables called *negative association*.

Definition 6 (Negative association [\[JDP83\]](#)). *Random variables $\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n$ are said to be negatively associated if, for every pair of disjoint subsets $A_1, A_2 \subseteq [n]$ and nondecreasing*

OCS $_{\mathcal{W}, \mathcal{T}}$

1. **Initialization:** For each $i \in L$, independently draw a $\mathbf{x}(i) \sim \mathcal{W}$ and initialize an index $k_1(i) \leftarrow 1$
2. Upon receiving $A_j = \{i_1, \dots, i_m\}$
 - (a) **Determine desired win probabilities:** For each $i \in A_j$, let $r_j(i)$ be its multiplicity in A_j . Compute the desired win probability

$$\mathbf{w}_j(i) := 1 - \prod_{\ell=0}^{r_j(i)-1} (1 - \mathbf{x}_{k_j(i)+\ell}) \quad (2)$$

- (b) **Choose a winner:** Use a (possibly randomized) tournament subroutine, \mathcal{T} , to select a single $i \in A_j$ as winner. This is done so that the probability each $i \in A_j$ wins is at least $\mathbf{w}_j(i)$.
- (c) **Increment counters:** For each $i \in L$, increment $k_{j+1}(i) \leftarrow k_j(i) + r_j(i)$

Figure 2: Our meta algorithm for constructing OCSes as a function of a win distribution, \mathcal{W} and tournament subroutine \mathcal{T}

functions f_1, f_2 ,

$$\text{Cov}[f_1(\mathbf{x}_{A_1}), f_2(\mathbf{x}_{A_2})] \leq 0$$

where \mathbf{x}_A is the vector $(\mathbf{x}_{i_1}, \mathbf{x}_{i_2}, \dots, \mathbf{x}_{i_{|A|}})$ when $A = \{i_1, i_2, \dots, i_{|A|}\}$.

Proving this ends up being essential to showing our OCS has *any* guarantees, even those matching the trivial fully independent OCS.

2. A *quantitative* and *short-distanced* negative correlation. We want the values of \mathbf{x}_j , and $\mathbf{x}_{j'}$ to be strongly negatively correlated whenever j is “close to” j' . The magnitude of this negative correlation directly ties into the probability that i is selected in consecutive time steps and therefore the quality of our OCS.

We prove a meta theorem on the quality of the algorithm in [Figure 2](#) as a function of the win distribution.

Theorem 4 (Quality of our meta algorithm). *Let \mathcal{W} be a win-distribution with the following three properties.*

1. **Negatively associated:** For any $n \geq 1$ and $\mathbf{x} \sim \mathcal{W}$, the variables $\mathbf{x}_1, \dots, \mathbf{x}_n$ are negative associated.
2. **Shift-invariant:** For any $n \geq 1$ and $\mathbf{x} \sim \mathcal{W}$, the distribution of $(\mathbf{x}_1, \mathbf{x}_2, \dots)$ is identical to that of $(\mathbf{x}_n, \mathbf{x}_{n+1}, \dots)$.

3. **Sufficiently small:** For each $r \in [m - 1]$ and $t \in [0, 1]$, defining $\mathbf{w}^{(r)}$ to be the random variable representing the distribution of the desired win probability for a vertex of multiplicity r ,

$$\mathbf{w}^{(r)} \sim \left(1 - \prod_{\ell=1}^r (1 - \mathbf{x}_\ell) \right) \quad \text{where} \quad \mathbf{x} \sim \mathcal{W},$$

the following holds

$$\mathbb{E}_{\mathbf{x} \sim \mathcal{W}} \left[\left(\mathbf{w}^{(r)} - t \right)_+ \right] \leq 1 - t - \frac{m - r}{m} \cdot \left(1 - t^{\frac{m}{m-r}} \right) \quad (3)$$

where $(z)_+$ is short hand for $\max(z, 0)$.

Then, there exists some tournament subroutine \mathcal{T} for which $\text{OCS}_{\mathcal{W}, \mathcal{T}}$ is a (f, m) -discrete OCS for

$$F(n) := \mathbb{E}_{\mathbf{x} \sim \mathcal{W}} \left[\prod_{\ell=1}^n (1 - \mathbf{x}_\ell) \right]. \quad (4)$$

Each of the three criteria of [Theorem 4](#) are desirable for intuitive reasons. Let S_1, \dots, S_k be disjoint set of time steps. Negative association ensures that,

$$\Pr[i \text{ never selected in } \bigcup_{\ell=1}^k S_\ell] \leq \prod_{\ell=1}^k \Pr[i \text{ never selected in } S_\ell].$$

Shift-invariance ensures that the probability that i is not selected in a set of n consecutive appearances is independent of when those appearances occur. Finally, $\mathbf{w}^{(r)}$ being sufficiently small ensures that it's possible for a tournament subroutine to guarantee that the probability i wins a round with desired win probability w is at least w .

Our proof of [Theorem 4](#) leverages a number of tools from probability theory. We use many closure properties of negatively associated variables given in [\[JDP83\]](#) and also prove some additional ones. Later, we formalize what it means for one variable to be “larger and more spread out” than another using the *increasing convex order*. We do so to take advantage of a powerful result by Müller and Rüschendorf [\[MR01\]](#) which generalizes a classic result of Strassen [\[Str65\]](#). That result guarantees the existence of a particular coupling (\mathbf{a}, \mathbf{b}) whenever \mathbf{b} is “larger and more spread out” than \mathbf{a} .

4 Organization of the remainder of this paper

[Section 5](#) compares our results to concurrent works and [Section 6](#) summarizes our notation. The remainder of this paper is devoted to proving our main theorem:

Theorem 5. *There is a $\Gamma = 0.5368$ -competitive algorithm for edge-weighted online bipartite matching.*

The sections are ordered as successive reductions from edge-weighted online bipartite matching with some proofs deferred to the appendix.

- In [Section 7](#), we give the formal version of [Definition 2](#), specifying the requirements a continuous OCS has to meet to be useful for edge-weighted online bipartite matching. We also state and prove the formal version of [Theorem 1](#) mapping that definition to a competitive ratio.
- In [Section 8](#), we show to construct a continuous OCS from a discrete OCS and then give the formal version of [Theorem 2](#) mapping a discrete OCS to a competitive ratio. We also give the proof of [Theorem 3](#) and of [Theorem 5](#) as a consequence of the discrete OCS construction in later sections.
- In [Section 9](#), we prove [Theorem 4](#) which gives a discrete OCS from a win distribution.
- In [Section 10](#), we construct a win distribution meeting the requirements of [Theorem 4](#) and optimize its parameters to achieve a 0.5368-competitive algorithm.
- Finally, in [Section 11](#), we provide a short proof of [Lemma 2.2](#) upper bounding the competitive ratio achievable by the OCS definition of [\[FHTZ20\]](#).

5 Comparison to the concurrent work

5.1 The work of Gao et al.

Concurrently and independently, Gao et al. also explored the use of multiway OCSes in the construction of improved online matching algorithms [\[GHH⁺21\]](#). The bulk of their work focuses on *vertex-weighted* online bipartite matching. This is a special case of *edge-weighted* online bipartite matching where we assume that all edges to each offline vertex have the same weight. While an algorithm achieving the optimal competitive ratio of $(1 - \frac{1}{e}) \approx 0.632$ is known in this setting [\[AGKM11\]](#), Gao et al. are able to show that OCSes can also be used to tackle this problem and achieve a competitive ratio of 0.593.

They do so by developing a definition of multiway OCSes similar to our definition of continuous OCSes. In the vertex-weighted setting we only care about whether an element i is selected at least once at some time step and do not care about which time step it is selected in. Their definition for a “multi-way semi-OCS” is equivalent to [Definition 2](#) where we take $k = 1$ and S_1 the set containing all time steps. For vertex-weighted online bipartite matching, they derive the same formula for the competitive ratio as we do in [Theorem 1](#) using “multi-way semi-OCSes.” A similar result could be derived using our algorithm by never executing [Equation \(10\)](#) (since we only care about whether a vertex is matched, not when it is matched). The resulting algorithm and analysis simplifies.

We stress that extending multiway OCSes to the historically challenging edge-weighted setting requires overcoming additional hurdles.

1. The informal definition of [Definition 2](#) isn’t strong enough to derive [Theorem 1](#). We needed to identify an additional constraint on the OCS (see “partially ignore small gaps” of [Definition 12](#)) that is both feasible to construct and sufficient to developing an algorithm for edge-weighted online bipartite matching.
2. Constructing a multiway OCS is more challenging than a “multi-way semi-OCS” and we need to develop a new framework to do so. For example, in the construction of Gao et al., the

probability a vertex is selected in a late time step can be made arbitrarily small (since they only need each vertex to be matched once), and so would not meet our definition.

Gao et al. also contribute to the *edge-weighted* setting. In this setting, they give 0.519-competitive algorithm by constructing an improved 2-way OCS. Their result improves upon the competitive ratio of Fahrbach et al. [FHTZ20] though isn't as large as ours. Finally, they prove that no γ -OCS exists for $\gamma > \frac{1}{4}$, which is stronger than the $\frac{1}{3}$ bound we show in [Lemma 2.2](#). For our purposes, the conclusion is the same: Multiway OCSes are strictly more powerful than two-way OCSes for edge-weighted online bipartite matching.

5.2 The work of Shin and An

Also concurrently and independently, Shin and An created a 3-discrete OCS and applied it to edge-weighted online bipartite matching [SA21]. They construct their 3-discrete OCS by running two parallel instances of a γ -OCS (which is 2-discrete). Upon receiving a set of three elements, two are chosen uniformly at random to be passed to one of the γ -OCSes. The winner and the third element are passed into the other γ -OCS to decide the overall winner of the round.

Then, they show how to use this 3-discrete OCS for edge-weighted online bipartite matching. To do so, they extend the analysis and factor-revealing LP of Fahrbach et al. and derive a competitive ratio of 0.5093. We note that plugging in the parameters of their 3-discrete OCS into [Theorem 2](#) gives the same competitive ratio up to the four digits of accuracy provided in their paper, though with a completely different derivation than theirs.

6 Preliminaries

For brevity, we use *decreasing* in place of *nondecreasing* and *increasing* in place of *nondecreasing*. The notation $(x)_+$ is used as shorthand for $\max(x, 0)$.

We use **boldface** (e.g. $\mathbf{x} \sim \mathcal{D}$) to denote random variables. The notation $\mathbf{a} \stackrel{d}{=} \mathbf{b}$ signifies that \mathbf{a} and \mathbf{b} have the same distribution. The following distributions are of interest.

1. The uniform distribution $\mathbf{x} \sim \text{Uniform}(0, 1)$, gives a continuous \mathbf{x} equally likely to be anywhere on the interval $[0, 1]$.
2. The Bernoulli distribution, $\mathbf{x} \sim \text{Ber}(p)$, gives a discrete \mathbf{x} that is 1 with probability p and 0 otherwise.
3. The geometric distribution, $\mathbf{x} \sim \text{Geo}(p)$ gives $i - 1$, where i is the smallest index such that $\mathbf{y}_i = 1$ for $\mathbf{y}_1, \mathbf{y}_2, \dots \stackrel{\text{iid}}{\sim} \text{Ber}(p)$.
4. The Poisson distribution, $\mathbf{x} \sim \text{Poi}(\lambda)$ gives \mathbf{x} defined as the following limit

$$\mathbf{x} := \lim_{n \rightarrow \infty} \left(\sum_{i=1}^n \mathbf{x}_i \quad \text{where} \quad \mathbf{x}_i \stackrel{\text{iid}}{\sim} \text{Ber}(\lambda/n) \right).$$

5. The categorical distribution, $\mathbf{i} \sim \text{Categorical}(p_1, \dots, p_m)$ where $\sum_{j=1}^m p_j = 1$, gives a discrete \mathbf{i} where $\mathbf{i} = i$ with probability p_i .

We'll need various ways to compare distributions. The simplest is *stochastic dominance*.

Definition 7 (Stochastic dominance). *Let \mathbf{a} and \mathbf{b} be random variables over \mathbb{R} . We say that \mathbf{a} stochastically dominates \mathbf{b} , denoted $\mathbf{a} \succeq \mathbf{b}$ if, for any $t \in \mathbb{R}$,*

$$\Pr[\mathbf{a} \leq t] \leq \Pr[\mathbf{b} \leq t]$$

Definition 8 (Stochastically increasing/decreasing functions). *A stochastic function \mathbf{f} from \mathbb{R}^n to a distribution over \mathbb{R} is said to be increasing in coordinate $i \in [n]$ if for any $x, y \in \mathbb{R}^n$ where $x_j = y_j$ for each $j \neq i$ and $x_i \geq y_i$, $\mathbf{f}(x)$ stochastically dominates $\mathbf{f}(y)$. It is said to be stochastically decreasing in coordinate i if the $\mathbf{f}(y)$ stochastically dominates $\mathbf{f}(x)$ under the same conditions. A function is stochastically increasing (or decreasing) if it is stochastically increasing (or decreasing) in all of its coordinates.*

In some proofs, we'll need for one distribution to be "less spread out" than another. We use the following two notions. For a more thorough overview, see [SS07].

Definition 9 (Convex order). *Let \mathbf{a}, \mathbf{b} be two univariate random variables such that,*

$$\mathbb{E}[\phi(\mathbf{a})] \leq \mathbb{E}[\phi(\mathbf{b})]$$

for all convex functions $\phi : \mathbb{R} \rightarrow \mathbb{R}$ for which the above expectations exist. Then, we say \mathbf{a} is smaller than \mathbf{b} in the convex order and denote this,

$$\mathbf{a} \preceq_{\text{cx}} \mathbf{b}.$$

Similar to the above, but where we restrict ϕ to be increasing,

Definition 10 (Increasing convex order). *Let \mathbf{a}, \mathbf{b} be two univariate random variables such that,*

$$\mathbb{E}[\phi(\mathbf{a})] \leq \mathbb{E}[\phi(\mathbf{b})]$$

for all increasing convex functions $\phi : \mathbb{R} \rightarrow \mathbb{R}$ for which the above expectations exist. Then, we say \mathbf{a} is smaller than \mathbf{b} in the increasing convex order and denote this,

$$\mathbf{a} \preceq_{\text{icx}} \mathbf{b}.$$

Finally, we'll use *discrete derivatives*.

Definition 11 (Discrete derivative). *For any function $F : \mathbb{Z} \rightarrow \mathbb{R}$, the discrete derivative of F is defined as*

$$(\Delta F)(n) := F(n+1) - F(n).$$

We use the notation $\Delta^{(\ell)}$ to denote the discrete derivative operator composed ℓ times.

7 Continuous OCS: Formal definition and theorems

In this section we formalize our definition of the parameters of a continuous OCS and the relation between those parameters and the competitive ratio. Recall in [Definition 2](#), we guarantee the probability i isn't picked in two consecutive sequences of time steps, S_1 and S_2 , is at most

$$f(\text{weight}_i(S_1)) \cdot f(\text{weight}_i(S_2)). \quad (5)$$

where $\text{weight}_i(S) = \sum_{j \in S} (p^{(j)})_i$. On the other hand, if S_1 and S_2 are actually consecutive, then the probability i isn't picked is at most

$$f(\text{weight}_i(S_1) + \text{weight}_i(S_2)). \quad (6)$$

Good OCSes will have $f(x + y) < f(x) \cdot f(y)$ and so [Equation \(6\)](#) gives a stronger guarantee than [Equation \(5\)](#). Now, suppose that S_1 and S_2 are ‘‘almost’’ consecutive. For $S_1 = \{j_1, \dots, j_2\}$ and $S_2 = \{j_3, \dots, j_4\}$, perhaps its the case that $j_3 = j_2 + 2$. Then, intuitively, we should expect a guarantee close to that of [Equation \(6\)](#) to hold.

Let G be the time steps in the ‘‘gap’’, meaning $G := \{j_2 + 1, \dots, j_3 - 1\}$. For some rate parameter r and $w := \prod_{j \in G} (1 - p_i^{(j)})^r$, we will guarantee that the probability i is not selected in S_1 or S_2 is at most

$$w \cdot f(\text{weight}_i(S_1) + \text{weight}_i(S_2)) + (1 - w) \cdot f(\text{weight}_i(S_1)) \cdot f(\text{weight}_i(S_2))$$

Essentially, when determining the probability i is picked, we pretend the gap G doesn't exist with probability w . This guarantee smoothly interpolates between [Equations \(5\)](#) and [\(6\)](#) as S_1 and S_2 grow further apart. We generalize this idea to more than two sets in the following definition.

Definition 12 (Parameters of a continuous OCS, formal version of [Definition 2](#)). *For any rate $r \geq 0$ and function $f : \mathbb{R}_{\geq 0} \rightarrow [0, 1]$, an (f, r) -continuous OCS is a continuous OCS with the following guarantee. For any element $i \in L$ and set of time steps S ,*

$$\Pr[\mathbf{i}_j \neq i \text{ for every } j \in S] \leq \mathcal{F}(S, p)$$

where p is the vector whose j^{th} entry is $(p^{(j)})_i$ and \mathcal{F} satisfies

1. **Consecutive:** If $S = \{j_1, \dots, j_2\}$, then

$$\mathcal{F}(S, p) \leq f\left(\sum_{j \in S} p_j\right)$$

2. **Partially ignore small gaps:** For any disjoint (not necessarily consecutive) $S_1, S_2 \subseteq \mathbb{N}$ satisfying

$$j_1 < j_2 \text{ for each } j_1 \in S_1, j_2 \in S_2$$

Let j^* be some time step between j_1, j_2 for every $j_1 \in S_1, j_2 \in S_2$. For

$$\begin{aligned} (\text{del}_{j^*}(p))_j &:= \begin{cases} p_j & j < j^* \\ p_{j+1} & j \geq j^* \end{cases} \\ \text{del}_{j^*}(S_1 \cup S_2) &:= \{j \text{ for each } j \in S_1\} \cup \{j - 1 \text{ for each } j \in S_2\} \end{aligned}$$

Then, for $w = (1 - p_{j^*})^r$,

$$\mathcal{F}(S_1 \cup S_2, p) \leq w \cdot \mathcal{F}(\text{del}_{j^*}(S_1 \cup S_2), \text{del}_{j^*}(p)) + (1 - w) \cdot \mathcal{F}(S_1, p) \cdot \mathcal{F}(S_2, p) \quad (7)$$

Smaller r corresponds to ignoring larger gaps, making the OCS anticorrelated over larger distances. Instead of setting $w = (1 - p_{j^*})^r$, we could have instead set $w = e^{-p_{j^*} \cdot r}$, $w = (1 - r \cdot p_{j^*})_+$, or other functions where $-r \leq \frac{dw}{dp_{j^*}} \leq 0$, and our proof of [Theorem 6](#) would still work. The choice we make here plays well with the proof of [Lemma 8.1](#).

We next give a formal version of [Theorem 1](#), with the main difference being that r can't be too large.

Theorem 6 (Formal version of [Theorem 1](#)). *For any thrice differentiable convex $f : \mathbb{R}_{\geq 0} \rightarrow [0, 1]$ where f and f' are log-concave, meaning*

$$\begin{aligned} f''(x) \cdot f(x) &\leq (f'(x))^2 && \text{for all } x \geq 0 \\ f'''(x) \cdot f'(x) &\leq (f''(x))^2 && \text{for all } x \geq 0 \end{aligned}$$

let,

$$\Gamma := 1 - \int_0^\infty e^{-t} f(t) dt. \quad (8)$$

For any r satisfying

$$r \leq \frac{-f'(0) - \Gamma}{\Gamma + (1 - \Gamma) \cdot f'(0)} \quad (9)$$

If there is a (f, r) -continuous OCS, then there is a Γ -competitive algorithm for edge-weighted online bipartite matching.

In most cases, $f'(0) = -1$ as OCSes of interest have f between $f_{\text{trivial}} = e^{-x}$ and $f_{\text{perfect}} = (1 - x)_+$. When $f'(0) = 1$, [Equation \(9\)](#) imposes an upper bound on r of $\frac{1 - \Gamma}{2\Gamma - 1}$, which is a decreasing function of Γ on the interval $(0.5, 1)$. Hence, the larger a competitive ratio we want, the smaller r has to be, meaning the OCS must be anticorrelated over larger distances.

The remainder of this section is devoted to proving [Theorem 6](#). The algorithm uses an online primal-dual approach.

7.1 Primal dual formulation

We begin with the standard primal linear program for (offline) edge-weighted bipartite matching. Let w_{ij} be the weight of the edge between $i \in L$ to $j \in R$ where $w_{ij} = 0$ indicate no such edge exists and x_{ij} be the probability that (i, j) is the heaviest edge matched to vertex i .

$$\begin{aligned} \max \quad & \sum_{i \in L} \sum_{j \in R} w_{ij} x_{ij} \\ \text{s.t.} \quad & \sum_{j \in R} x_{ij} \leq 1 && \forall i \in L \\ & \sum_{i \in L} x_{ij} \leq 1 && \forall j \in R \\ & x_{ij} \geq 0 && \forall i \in L, j \in R \end{aligned}$$

This gives rise to a dual linear program.

$$\begin{aligned}
\min \quad & \sum_{i \in L} \alpha_i + \sum_{j \in R} \beta_j \\
\text{s.t.} \quad & \alpha_i + \beta_j \geq w_{ij} && \forall i \in L, j \in R \\
& \alpha_i \geq 0 && \forall i \in L \\
& \beta_j \geq 0 && \forall j \in R
\end{aligned}$$

We use P to denote the primal objective and D to denote the dual objective. Our online algorithm will simultaneously maintain a matching for the graph and a dual assignment that will meet the following two criteria.

Fact 7.1 (Lemma 2 from [FHTZ20]). *Suppose an algorithm maintains primal and dual assignments such that, for some $0 \leq \Gamma \leq 1$, the following conditions hold at all time steps.*

1. **Approximate dual feasibility:** For any $i \in L$ and $j \in R$

$$\alpha_i + \beta_j \geq \Gamma \cdot w_{ij}$$

2. **Reverse weak duality:** The objectives of the primal and dual assignments satisfy $P \geq D$.

Then the algorithm is Γ -competitive.

Inspired by [DHK⁺16, FHTZ20], we do not directly maintain the primal variables x_{ij} . Instead, for each $i \in L$ and weight-level $w \geq 0$, we define

$$y_i(w) := \Pr[i \text{ not matched to a vertex of weight at least } w]$$

In contrast, [DHK⁺16, FHTZ20] defines their $\overline{y_i(w)}$ to be $\overline{y_i(w)} = 1 - y_i(w)$. We find our “reversed” definition simplifies later analysis. The primal objective is then

$$P = \sum_{i \in L} \int_0^\infty (1 - y_i(w)) dw$$

Similarly, we introduce and maintain new variables $\alpha_i(w)$ for each $i \in L$ and weight-level $w \geq 0$ and then set the value of α_i to

$$\alpha_i = \int_0^\infty (\Gamma - \alpha_i(w)) dw.$$

Remark 1. Note that $y_i(w)$ and $\alpha_i(w)$ are both step functions with at most one change per unique value in the set $\{w_{ij} \mid j \in R\}$. Hence, they can be stored efficiently.

7.2 High level algorithmic overview

Fix some (f, r) -continuous OCS. At each time step $j \in [T]$, the values w_{ij} are revealed for every $i \in L$. As a function of these weights and the state of the primal and dual variables, the algorithm will determine a probability vector $p^{(j)}$ to pass into the OCS. Then, it updates the primal variables to account for the probability each vertex is matched to j and the dual variables to maintain the

conditions dual feasibility and reverse weak duality. The OCS selects a winner in L based on the $p^{(j)}$ it is passed in, and that winner is matched to vertex j . We give pseudocode for the entire algorithm at the end of this section in [Figure 3](#).

We wish to emphasize that the primal-dual algorithm does not “know” which offline vertex j is matched too. Instead, it only maintains the probability that each i is matched to a vertex at each weight-level. This separates the algorithm into a deterministic component, the primal-dual portion, and a randomized component, the OCS.

The remainder of this section is divided as follows. In [Section 7.3](#) and [Section 7.4](#), we define how the primal and dual variables respectively are updated given $p^{(j)}$. In particular, β_j is set to exactly the minimum value so that approximate dual feasibility holds. In [Section 7.5](#) we will discuss how $p^{(j)}$ is determined. In [Section 7.6](#), we will prove that reverse weak duality holds, completing the proof that our algorithm is Γ -competitive.

Throughout this section, $f : \mathbb{R}_{\geq 0} \rightarrow [0, 1]$ and $r \in \mathbb{R}^+$ refer to the parameters of our OCS.

7.3 Primal updates

For each $i \in L$ and weight-level $w \geq 0$, we maintain two random variables initialized as follows.

$$\begin{aligned} \mathbf{x}_i(w) &\stackrel{\text{init}}{\leftarrow} 0 && \text{with probability } 1 \\ \mathbf{c}_i(w) &\stackrel{\text{init}}{\leftarrow} 1 && \text{with probability } 1 \end{aligned}$$

Our algorithm maintains the *distribution* over $(\mathbf{x}_i(w), \mathbf{c}_i(w))$ rather than just a single element of it. See [Remark 2](#) for the computational complexity of maintaining this distribution.

At time step j , our algorithm will pass a vector $p^{(j)}$ into the OCS. After doing so, it updates, for each $w \leq w_{ij}$,

$$\mathbf{x}_i(w) \leftarrow \mathbf{x}_i(w) + (p^{(j)})_i$$

For each $w > w_{ij}$, with probability $1 - (1 - (p^{(j)})_i)^r$ we set

$$\begin{aligned} \mathbf{c}_i(w) &\leftarrow \mathbf{c}_i(w) \cdot f(\mathbf{x}_i(w)) \\ \mathbf{x}_i(w) &\leftarrow 0 \end{aligned} \tag{10}$$

Otherwise, we leave $\mathbf{c}_i(w)$ and $\mathbf{x}_i(w)$ unchanged. This is the step that makes $\mathbf{c}_i(w)$ and $\mathbf{x}_i(w)$ random quantities rather than fixed.

At all steps, we maintain that

$$y_i(w) = \mathbb{E}[\mathbf{c}_i(w) \cdot f(\mathbf{x}_i(w))]$$

Note that these updates correspond to [Definition 12](#). For $S = \{j \in R \mid w_{ij} \geq w\}$, that definition guarantees the probability that i is not matched to a vertex in S is at most $y_i(w)$.

Remark 2. The full support of $(\mathbf{x}_i(w), \mathbf{c}_i(w))$ can be exponentially large, and so inefficient to maintain. However, our updates never need full access to the distribution over $(\mathbf{x}_i(w), \mathbf{c}_i(w))$. It is sufficient to maintain a distribution over the possible values of $\mathbf{x}_i(w)$, and for each possible value for $\mathbf{x}_i(w) = x$, the quantity $\mathbb{E}[\mathbf{c}_i(w) \mid \mathbf{x}_i(w) = x]$. Since there are only $|R|$ many possible values for $\mathbf{x}_i(w)$, that information can be stored and operated on efficiently.

7.4 Dual updates

Our dual updates depend on some function $a : \mathbb{R}_{\geq 0} \rightarrow [0, \Gamma]$ with the properties in the below Lemma.

Lemma 7.2. *Let $f : \mathbb{R}_{\geq 0} \rightarrow [0, 1]$ be thrice differentiable, convex, log concave, have a log concave derivative, and satisfy $f(0) = 1$. Define $a : \mathbb{R}_{\geq 0} \rightarrow [0, \Gamma]$ as*

$$a(x) := f(x) - \int_0^\infty e^{-t} f(t+x) dt. \quad (11)$$

For r satisfying Equation (9) and Γ defined in Equation (8), a satisfies the below properties.

$$a'(x) = f'(x) + a(x) \quad \forall x \geq 0 \quad (12)$$

$$a'(x) \leq 0 \quad \forall x \geq 0 \quad (13)$$

$$a(0) = \Gamma \quad (14)$$

$$f(x) \cdot a(0) = f(x) \cdot \Gamma \leq a(x) \quad \forall x \geq 0 \quad (15)$$

$$r \cdot (a(x) - f(x) \cdot \Gamma) \leq \Gamma - a(x) \quad \forall x \geq 0 \quad (16)$$

We prove Lemma 7.2 in Appendix B. For the remainder of this section, let a be as defined in Equation (11).

For any $i \in L$ and weight-level $w \geq 0$, we maintain that

$$\alpha_i(w) = \mathbb{E} \left[\mathbf{c}_i(w) \cdot a(\mathbf{x}_i(w)) \right],$$

where $\mathbf{c}_i(w)$, $\mathbf{x}_i(w)$ are the same as in the primal updates. After α_i is updated at time step j , we set β_j to the minimum value so that approximate dual feasibility holds. Specifically,

$$\beta_j = \max \left(0, \max_{i \in L} \Gamma \cdot w_{ij} - \alpha_i \right) \quad (17)$$

For the above setting of β_j , dual feasibility will hold for w_{ij} at time step j . It will remain satisfied as long α_i never decreases in future time steps. At any time step j' , if $(p^{(j')})_i = 0$, then α_i is not changed at that time step. The following proposition guarantees that if $(p^{(j')})_i > 0$, then α_i can only increase, which implies dual feasibility remains satisfied.

Proposition 7.3. *Fix some time step j . For each $i \in L$, $w \geq 0$ and $p \in [0, 1]$, let $\alpha_i[p]$ and $\alpha_i(w)[p]$ denote the value that would be assigned to α_i and $\alpha_i(w)$ respectively at the end of this time step if $(p^{(j)})_i = p$. Then,*

1. $\alpha_i[p]$ is a continuous and increasing function of p .
2. $\alpha_i(w)[p]$ is a continuous and decreasing function of p .

Proof. Recall that

$$\alpha_i := \int_0^\infty (\Gamma - \alpha_i(w)) dw,$$

so it is sufficient to prove that $\alpha_i(w)[p]$ is a continuous and decreasing function of p . First, consider the case where $w \leq w_{ij}$. Then, using $\mathbf{x}_i(w), \mathbf{c}_i(w)$ as their values *before* the updates from round j ,

$$\frac{d}{dp}\alpha_i(w)[p] = \mathbb{E}[\mathbf{c}_i(w) \cdot a'(\mathbf{x}_i(w) + p)].$$

This derivative exists and is at most 0 by [Equation \(13\)](#), so $\alpha_i(w)[0]$ is continuous and decreasing.

Next, consider the case where $w > w_{ij}$. Then,

$$\begin{aligned} \frac{d}{dp}\alpha_i(w)[p] &= \frac{d}{dp}(1 - (1 - p)^r) \cdot \mathbb{E} \left[\mathbf{c}_i(w) \cdot (f(\mathbf{x}_i(w)) \cdot a(0) - a(\mathbf{x}_i(w))) \right] \\ &= r \cdot (1 - p)^{r-1} \cdot \mathbb{E} \left[\mathbf{c}_i(w) \cdot (f(\mathbf{x}_i(w)) \cdot a(0) - a(\mathbf{x}_i(w))) \right] \\ &\leq 0 \end{aligned} \tag{Equation (15)}$$

Once again, the derivative exists and is less than 0. \square

7.5 Determining the probability vector

At time step j , our algorithm receives w_{ij} for all $i \in L$. It then sets $p^{(j)}$ to a vector satisfying [Lemma 7.4](#).

Note that the $p^{(j)}$ given by [Lemma 7.4](#) can have a sum of elements that is less than 1, but [Definition 1](#) requires the sum of the elements to be 1. To remedy this, we just create a dummy vertex and assign any excess probability to it.

Lemma 7.4. *It is always possible to choose $p^{(j)}$ so that, after the updates corresponding to $p^{(j)}$,*

$$\beta_j = \sum_{i \in L} (p^{(j)})_i \cdot (\Gamma \cdot w_{ij} - \alpha_i) \tag{18}$$

Proof. Let $\alpha_i[p]$ be as defined in [Proposition 7.3](#). For each $i \in L$ and $\beta \geq 0$, let $p_i[\beta]$ be the lowest possible value of $p \geq 0$ such that $\Gamma \cdot w_{ij} - \alpha_i[p] \leq \beta$. We note two properties about this function.

1. $p_i[\beta]$ is a continuous and decreasing function of β . This follows from [Proposition 7.3](#)'s guarantee that $\alpha_i[p]$ is a continuous and increasing function of p .
2. For any $\beta \geq 0$, if $p_i[\beta] \geq 0$, then $\Gamma \cdot w_{ij} - \alpha_i[p] = \beta$. This follows from the continuity of $\alpha_i[p]$.

First, consider the case where $\sum_{i \in L} p_i[0] \leq 1$. In this case, set $(p^{(j)})_i = p_i[0]$ for each $i \in L$. Then, β_j will be set to 0 in [Equation \(17\)](#) and

$$\begin{aligned} &\sum_{i \in L} (p^{(j)})_i \cdot (\Gamma \cdot w_{ij} - \alpha_i) \\ &= \sum_{i \in L} (p^{(j)})_i \cdot (0) \quad (p_i[\beta] \geq 0 \implies \Gamma \cdot w_{ij} - \alpha_i[p] = \beta \text{ applied with } \beta = 0) \\ &= 0 = \beta_j. \end{aligned}$$

Otherwise, $\sum_{i \in L} p_i[0] > 1$. Note that $\lim_{\beta \rightarrow \infty} \sum_{i \in L} p_i[\beta] = 0$. By the intermediate value theorem applied to $\beta \mapsto \sum_{i \in L} p_i[\beta]$, there is some choice β^* for which $\sum_{i \in L} p_i[\beta^*] = 1$.

In this case, set $(p^{(j)})_i = p_i[\beta^*]$ for each $i \in L$. Then, β_j will be set to β^* in Equation (17) and

$$\begin{aligned}
& \sum_{i \in L} (p^{(j)})_i \cdot (\Gamma \cdot w_{ij} - \alpha_i) \\
&= \sum_{i \in L} (p_i^{(j)}) \cdot \beta^* && (p_i[\beta] \geq 0 \implies w_{ij} - \alpha_i[p] = \beta \text{ applied with } \beta = \beta^*) \\
&= 1 \cdot \beta^* && (\sum_{i \in L} p_i[\beta^*] = 1) \\
&= \beta^* = \beta_j.
\end{aligned}$$

□

7.6 Reverse weak duality

Since we already have that approximate dual feasibility holds, we only need to prove that reverse weak duality holds to complete the proof that this algorithm is Γ -competitive. At the start of our algorithm we have that $P = D = 0$. Let $\Delta_j P$ and $\Delta_j D$ be the amount the primal and dual respectively increase after the update corresponding to $j \in R$. We will prove that $\Delta_j P \geq \Delta_j D$ which guarantees reverse weak duality holds.

Lemma 7.5. *For any $j \in R$, if $p^{(j)}$ is set as in Lemma 7.4, then $\Delta_j P \geq \Delta_j D$.*

Proof. For each primal and dual variable, we use a superscript of (j) to refer to its value after the j^{th} update. Similarly, a superscript of $(j-1)$ refers to its value before the update. For any variable var, we use the shorthand $\Delta_j \text{var}$ to refer $\text{var}^{(j)} - \text{var}^{(j-1)}$.

$$\begin{aligned}
\Delta_j P - \Delta_j D &= \left(- \sum_{i \in L} \int_0^\infty \Delta_j y_i(w) dw \right) - \left(- \sum_{i \in L} \int_0^\infty \Delta_j \alpha_i(w) dw + \beta_j \right) \\
&= \sum_{i \in L} \left(- \int_0^\infty (\Delta_j y_i(w) - \Delta_j \alpha_i(w)) dw \right) - \left(\sum_{i \in L} (p^{(j)})_i \cdot (\Gamma \cdot w_{ij} - \alpha_i^{(j)}) \right)
\end{aligned}$$

Recall that $\alpha_i = \int_0^\infty (\Gamma - \alpha_i(w)) dw$. Therefore,

$$\begin{aligned}
\Gamma \cdot w_{ij} - \alpha_i^{(j)} &= \left(\int_0^{w_{ij}} \Gamma \cdot dw \right) - \left(\int_0^\infty (\Gamma - \alpha_i^{(j)}(w)) dw \right) \\
&= \left(\int_0^{w_{ij}} \alpha_i^{(j)}(w) dw \right) - \left(\int_{w_{ij}}^\infty (\Gamma - \alpha_i^{(j)}(w)) dw \right)
\end{aligned}$$

Combining the above equations, we have that

$$\begin{aligned}
\Delta_j P - \Delta_j D &= \sum_{i \in L} \left(\int_0^{w_{ij}} \left(-\Delta_j y_i(w) + \Delta_j \alpha_i(w) - (p^{(j)})_i \cdot \alpha_i^{(j)}(w) \right) dw \right) \\
&\quad + \sum_{i \in L} \left(\int_{w_{ij}}^\infty \left(-\Delta_j y_i(w) + \Delta_j \alpha_i(w) + (p^{(j)})_i \cdot (\Gamma - \alpha_i^{(j)}(w)) \right) dw \right)
\end{aligned}$$

We will show that for any choice of $w \geq 0$ and $i \in L$, the corresponding term in the above equation is positive. This implies that $\Delta_j P - \Delta_j D$ is also positive.

Case 1: $w \leq w_{ij}$. We wish to show that $-\Delta_j y_i(w) + \Delta_j \alpha_i(w) - (p^{(j)})_i \cdot \alpha_i^{(j)}(w) \geq 0$. We expand each term separately.

$$\begin{aligned}\Delta_j y_i(w) &= y_i^{(j)}(w) - y_i^{(j-1)}(w) \\ &= \mathbb{E} \left[\mathbf{c}_i^{(j)}(w) \cdot (f(\mathbf{x}_i^{(j)}(w)) - f(\mathbf{x}_i^{(j)}(w) - (p^{(j)})_i)) \right] \\ &= \mathbb{E} \left[\int_{\mathbf{x}_i^{(j)}(w) - (p^{(j)})_i}^{\mathbf{x}_i^{(j)}(w)} \mathbf{c}_i^{(j)}(w) \cdot f'(x) dx \right]\end{aligned}$$

Similarly,

$$\begin{aligned}\Delta_j \alpha_i(w) &= \alpha_i^{(j)}(w) - \alpha_i^{(j-1)}(w) \\ &= \mathbb{E} \left[\mathbf{c}_i^{(j)}(w) \cdot (a(\mathbf{x}_i^{(j)}(w)) - a(\mathbf{x}_i^{(j)}(w) - (p^{(j)})_i)) \right] \\ &= \mathbb{E} \left[\int_{\mathbf{x}_i^{(j)}(w) - (p^{(j)})_i}^{\mathbf{x}_i^{(j)}(w)} \mathbf{c}_i^{(j)}(w) \cdot a'(x) dx \right]\end{aligned}$$

Finally,

$$\begin{aligned}(p^{(j)})_i \cdot \alpha_i^{(j)}(w) &= (p^{(j)})_i \cdot \mathbb{E} \left[\mathbf{c}_i^{(j)}(w) \cdot a(\mathbf{x}_i^{(j)}(w)) \right] \\ &= \int_{\mathbf{x}_i^{(j)}(w) - (p^{(j)})_i}^{\mathbf{x}_i^{(j)}(w)} \mathbb{E} \left[\mathbf{c}_i^{(j)}(w) \cdot a(\mathbf{x}_i^{(j)}(w)) \right] dx \\ &= \mathbb{E} \left[\int_{\mathbf{x}_i^{(j)}(w) - (p^{(j)})_i}^{\mathbf{x}_i^{(j)}(w)} \mathbf{c}_i^{(j)}(w) \cdot a(\mathbf{x}_i^{(j)}(w)) dx \right]\end{aligned}$$

Comparing the above three equations, it is enough to show that for any $x < x_i^{(j)}(w)$, that $-f'(x) + a'(x) - a(x_i^{(j)}) \geq 0$.

$$\begin{aligned}-f'(x) + a'(x) - a(x_i^{(j)}) &\geq -f'(x) + a'(x) - a(x) && (a \text{ is decreasing}) \\ &= 0 && \text{(Equation (12))}\end{aligned}$$

This completes the first case.

Case 2: $w > w_{ij}$. We wish to show that $-\Delta_j y_i(w) + \Delta_j \alpha_i(w) + (p^{(j)})_i \cdot (\Gamma - \alpha_i^{(j)}(w)) \geq 0$. For $w > w_{ij}$, $y_i(w)$ does not change. Hence, we only need to expand the other two terms. Recall that, for $w \geq w_{ij}$, we set $\mathbf{c}_i(w) \leftarrow \mathbf{c}_i(w) \cdot f(\mathbf{x}_i(w))$ and $\mathbf{x}_i(w) \leftarrow 0$ with probability $1 - (1 - (p^{(j)})_i)^r$. Otherwise, we leave the parameters unchanged. Expanding,

$$\Delta_j \alpha_i(w) = \mathbb{E} \left[(1 - (1 - (p^{(j)})_i)^r) \cdot \mathbf{c}_i^{(j-1)}(w) \cdot (f(\mathbf{x}_i^{(j-1)}(w)) \cdot a(0) - a(\mathbf{x}_i^{(j-1)}(w))) \right].$$

Recall from [Equation \(15\)](#) that $f(x) \cdot a(0) \leq a(x)$ for all $x \geq 0$. Therefore, we may replace the $1 - (1 - (p^{(j)})_i)^r$ factor with an upper bound on it, in this case $r \cdot (p^{(j)})_i$, and bound,

$$\Delta_j \alpha_i(w) \geq \mathbb{E} \left[r \cdot (p^{(j)})_i \cdot \mathbf{c}_i^{(j-1)}(w) \cdot (f(\mathbf{x}_i^{(j-1)}(w)) \cdot a(0) - a(\mathbf{x}_i^{(j-1)}(w))) \right].$$

Using the fact that $\alpha_i^{(j)}(w) \leq \alpha_i^{(j-1)}(w)$, which follows from [Proposition 7.3](#), we expand the other term of our desired inequality.

$$\begin{aligned} (p^{(j)})_i \cdot (\Gamma - \alpha_i^{(j)}(w)) &\geq (p^{(j)})_i \cdot (\Gamma - \alpha_i^{(j-1)}(w)) \\ &= (p^{(j)})_i \cdot \mathbb{E} \left[\Gamma - \mathbf{c}_i^{(j-1)}(w) \cdot a(\mathbf{x}_i^{(j-1)}(w)) \right] \end{aligned}$$

The desired inequality holds as long as, for all $x \geq 0$,

$$r \cdot (f(x) \cdot a(0) - a(x)) + (\Gamma - a(x)) \geq 0.$$

which is equivalent to [Equation \(16\)](#). This completes the proof of the second case and reverse weak duality. \square

8 From discrete to continuous OCSes

In this section, we show how to construct a continuous OCS from a discrete OCS. That will allow us to prove [Theorems 2](#) and [3](#).

Lemma 8.1. *If there is an (F, m) -discrete OCS for convex F , there is an $(f, r = m)$ -continuous OCS for*

$$f(x) := \mathbb{E}_{\mathbf{k} \sim \text{Poi}(\lambda = m \cdot x)} [F(\mathbf{k})] \tag{19}$$

The continuous OCS algorithm in [Lemma 8.1](#) is simple. When the continuous OCS receives the probability vector $p^{(j)}$, it sets the multiset $A_j = \{\mathbf{i}_1, \dots, \mathbf{i}_m\}$ where each \mathbf{i}_ℓ is sampled iid from $\text{Categorical}(p^{(j)})$. It passes this multiset into the (F, m) -discrete OCS, and then selects the same vertex selected by the discrete OCS. We include pseudocode for this reduction in [Figure 4](#).

Before analyzing this OCS, we'll need some well-known facts about the convex order (recall [Definition 9](#)).

Fact 8.2 (Theorem 3.A.4 of [\[SS07\]](#)). *For any univariate random variables \mathbf{a}, \mathbf{b} , if there exists a coupling $(\hat{\mathbf{a}}, \hat{\mathbf{b}})$ where $\mathbf{a} \stackrel{d}{=} \hat{\mathbf{a}}$ and $\mathbf{b} \stackrel{d}{=} \hat{\mathbf{b}}$ such that*

$$\mathbb{E}[\hat{\mathbf{b}} \mid \hat{\mathbf{a}}] = \hat{\mathbf{a}}$$

then $\mathbf{a} \preceq_{\text{cx}} \mathbf{b}$.

Fact 8.3 (Theorem 3.A.12(d) of [\[SS07\]](#)). *The convex order is closed under convolution: Let $\mathbf{a}_1, \dots, \mathbf{a}_k$ and $\mathbf{b}_1, \dots, \mathbf{b}_k$ each be independent univariate random variables. If $\mathbf{a}_\ell \preceq_{\text{cx}} \mathbf{b}_\ell$ for each $\ell = 1, \dots, m$, then*

$$\sum_{\ell=1}^m \mathbf{a}_\ell \preceq_{\text{cx}} \sum_{\ell=1}^m \mathbf{b}_\ell.$$

MATCHONLINE(L)

1. **Initialization:** For each $i \in L$, and $w \geq 0$, initialize the distributions $\mathbf{x}_i(w), \mathbf{c}_i(w)$ as

$$\begin{aligned}\mathbf{x}_i(w) &\stackrel{\text{init}}{\leftarrow} 0 && \text{with probability } 1 \\ \mathbf{c}_i(w) &\stackrel{\text{init}}{\leftarrow} 1 && \text{with probability } 1\end{aligned}$$

The algorithm maintains the *distribution* of possible $(\mathbf{x}_i(w), \mathbf{c}_i(w))$ over the randomness in [Step 3b](#).

2. **Invariants:** For each $i \in L$ and $w \geq 0$, maintain the primal variables, where the expectations are over the distribution of $(\mathbf{x}_i(w), \mathbf{c}_i(w))$,

$$y_i(w) = \mathbb{E}[\mathbf{c}_i(w) \cdot f(\mathbf{x}_i(w))]$$

and dual variables

$$\alpha_i(w) = \mathbb{E}[\mathbf{c}_i(w) \cdot a(\mathbf{x}_i(w))],$$

where a is the function from [Lemma 7.2](#).

3. Upon receiving w_{ij} for each $i \in L$
 - (a) **Determine input to OCS:** Using [Lemma 7.4](#), determine $p^{(j)}$ to satisfy [Equation \(18\)](#). This choice of $p^{(j)}$ is a function of the current state of the algorithm as well as w_{ij} for each $i \in L$.
 - (b) **Update distribution of $(\mathbf{x}_i(w), \mathbf{c}_i(w))$:** For each $i \in L$ and $w \leq w_{ij}$, update

$$\mathbf{x}_i(w) \leftarrow \mathbf{x}_i(w) + (p^{(j)})_i.$$

For each $w > w_{ij}$, with probability $1 - (1 - (p^{(j)})_i)^r$ set

$$\begin{aligned}\mathbf{c}_i(w) &\leftarrow \mathbf{c}_i(w) \cdot f(\mathbf{x}_i(w)) \\ \mathbf{x}_i(w) &\leftarrow 0\end{aligned}$$

Otherwise, leave $\mathbf{c}_i(w)$ and $\mathbf{x}_i(w)$ unchanged. This is the step that makes $\mathbf{c}_i(w)$ and $\mathbf{x}_i(w)$ random variables rather than fixed.

- (c) **Invariants:** Maintain the primal and dual variables as in [Step 2](#).
 - (d) **Match:** Pass $p^{(j)}$ into the OCS and match j to the winning vertex.

Figure 3: The algorithm for edge-weighted online bipartite matching. For computational efficiency concerns, see [Remarks 1](#) and [2](#).

ContinuousOCS(\mathcal{O}, m)

Upon receiving $p^{(j)}$

1. **Sample:** For each $k = 1, \dots, m$, sample independently $\mathbf{i}_k \sim \text{Categorical}(p^{(j)})$.
2. **Use discrete OCS:** Pass the multiset $A_j = \{\mathbf{i}_1, \dots, \mathbf{i}_m\}$ into \mathcal{O} and return the same winner as it.

Figure 4: A construction for a continuous OCS using black-box access to a m -discrete OCS \mathcal{O} .

The following proposition is a simple consequence of the above two facts.

Proposition 8.4. *For any $p_1, \dots, p_n \in [0, 1]$, let $\mathbf{x}_i \sim \text{Ber}(p_i)$ and $\mathbf{y} \sim \text{Poi}(\lambda = \sum_{i=1}^n p_i)$. Then,*

$$\sum_{i=1}^n \mathbf{x}_i \preceq_{\text{cx}} \mathbf{y}.$$

Proof. By [Fact 8.3](#), it is enough to prove that $\mathbf{x}_i \preceq_{\text{cx}} \mathbf{y}_i$ where $\mathbf{y}_i \sim \text{Poi}(\lambda = p_i)$. First, we note that \mathbf{y}_i is more likely to be 0 than \mathbf{x}_i .

$$\Pr[\mathbf{y}_i = 0] = \frac{p_i^0 \cdot e^{-p_i}}{0!} = e^{-p_i} \geq (1 - p_i) = \Pr[\mathbf{x}_i = 0].$$

Consider the following coupling: First draw $\hat{\mathbf{y}}_i \sim \text{Poi}(p_i)$. If $\hat{\mathbf{y}}_i = 0$, then with probability $(\Pr[\mathbf{x}_i = 0]) / (\Pr[\mathbf{y}_i = 0])$, set $\hat{\mathbf{x}}_i = 0$ as well. Otherwise, set $\hat{\mathbf{x}}_i = 1$.

We verify that this coupling meets the criteria of [Fact 8.2](#) which implies $\mathbf{x}_i \preceq_{\text{cx}} \mathbf{y}_i$. By design, we have that $\hat{\mathbf{y}}_i \stackrel{d}{=} \mathbf{y}_i$ and $\hat{\mathbf{x}}_i \stackrel{d}{=} \mathbf{x}_i$. If $\hat{\mathbf{x}}_i = 0$, then $\hat{\mathbf{y}}_i = 0$ as well, so

$$\mathbb{E}[\hat{\mathbf{y}}_i \mid \hat{\mathbf{x}}_i = 0] = 0$$

Then, using the law of total expectation, we have that

$$\mathbb{E}[\hat{\mathbf{y}}_i \mid \hat{\mathbf{x}}_i = 0] \cdot \Pr[\hat{\mathbf{x}}_i = 0] + \mathbb{E}[\hat{\mathbf{y}}_i \mid \hat{\mathbf{x}}_i = 1] \cdot \Pr[\hat{\mathbf{x}}_i = 1] = \mathbb{E}[\hat{\mathbf{y}}_i]$$

Substituting $\mathbb{E}[\hat{\mathbf{y}}_i \mid \hat{\mathbf{x}}_i = 0] = 0$, $\mathbb{E}[\hat{\mathbf{y}}_i] = p_i$, and $\Pr[\hat{\mathbf{x}}_i = 1] = p_i$, we have that

$$\mathbb{E}[\hat{\mathbf{y}}_i \mid \hat{\mathbf{x}}_i = 1] = 1.$$

Therefore, by [Fact 8.2](#), $\mathbf{x}_i \preceq_{\text{cx}} \mathbf{y}_i$. □

Now, we are ready to prove [Lemma 8.1](#).

Proof. Let \mathcal{A} be some (F, m) discrete OCS. We construct a continuous OCS \mathcal{B} as in [Figure 4](#): Whenever \mathcal{B} receives the probability vector $p^{(j)}$, it sets $\mathbf{A}_j = \{\mathbf{i}_1, \dots, \mathbf{i}_m\}$ where each \mathbf{i}_ℓ is sampled

iid from $\text{Categorical}(p^{(j)})$. It passes this multiset into \mathcal{A} , and selects the same winner that \mathcal{A} selected. By [Definition 4](#), for any disjoint consecutive sequences of time steps S_1, \dots, S_k ,

$$\Pr[i \text{ never selected in } S_1 \cup \dots \cup S_k] \leq \prod_{\ell=1}^k F(\mathbf{count}_i(S_\ell))$$

where $\mathbf{count}_i(S_\ell)$ is a random variable defined as

$$\mathbf{count}_i(S_\ell) := \sum_{j \in S_\ell} (\mathbf{r}_j(i)) \quad \text{where } \mathbf{r}_j(i) \text{ is the multiplicity of } i \text{ in } \mathbf{A}_j.$$

Based on how we defined \mathbf{A}_j , we have that $\mathbf{r}_j(i)$ is the sum of m independent variables distributed according to $\text{Ber}((p^{(j)})_i)$. Therefore, by [Proposition 8.4](#),

$$\mathbf{count}_i(S_\ell) \preceq_{\text{cx}} \mathbf{k}_\ell \quad \text{where } \mathbf{k}_\ell \sim \text{Poi} \left(m \cdot \sum_{j \in S_\ell} (p^{(j)})_i \right).$$

We are ready to show that \mathcal{B} meets the requirements of [Definition 12](#). First, we consider some consecutive $S = \{j_1, \dots, j_2\}$. Since F is convex,

$$\begin{aligned} \Pr[i \text{ not selected in } S] &= \mathbb{E}[F(\mathbf{count}_i(S))] \\ &\leq \mathbb{E}[F(\mathbf{k})] \quad \text{where } \mathbf{k} \sim \text{Poi} \left(m \cdot \sum_{j \in S} (p^{(j)})_i \right) \quad (\mathbf{count}_i(S) \preceq_{\text{cx}} \mathbf{k}) \\ &= f \left(\sum_{j \in S} (p^{(j)})_i \right) \quad (\text{Equation (19)}) \end{aligned}$$

For the second requirement, ‘‘partially ignore small gaps,’’ of [Definition 12](#), consider any disjoint $S_1, S_2 \subseteq \mathbb{N}$ where every element of S_1 precedes every element of S_2 . Define j^* as in [Definition 12](#). Then, the probability $\mathbf{r}_{j^*}(i) = 0$ is $w = (1 - (p^{(j^*)})_i)^m$. If $\mathbf{r}_{j^*}(i) = 0$ then the probability i is selected in $S_1 \cup S_2$ is the same as if the time step j^* didn’t exist (since \mathcal{A} won’t see i in it). Hence, [Equation \(7\)](#) holds. \square

In order to give a formal version of [Theorem 2](#) we’ll need to know the derivatives of f . There is a convenient relation between the *discrete derivatives* of F and the *derivatives* of f .

Lemma 8.5. *Choose any $F : \mathbb{Z}_{\geq 0} \rightarrow \mathbb{R}$, $m \in \mathbb{Z}$ and let $f : \mathbb{R}_{\geq 0} \rightarrow \mathbb{R}$ be*

$$f(x) = \mathbb{E}_{\mathbf{k} \sim \text{Poi}(mx)} [F(\mathbf{k})].$$

Then, for any $\ell \in \mathbb{N}$ and $x \in \mathbb{R}_{\geq 0}$, the derivatives satisfy the relation

$$f^{(\ell)}(x) = m^\ell \cdot \mathbb{E}_{\mathbf{k} \sim \text{Poi}(mx)} [\Delta^{(\ell)} F(\mathbf{k})].$$

Proof. It is enough to prove the case where $\ell = 1$, as larger ℓ follow by induction. The remainder of this proof is algebraic manipulations. We expand

$$\begin{aligned} f'(x) &= \frac{d}{dx} \left(\sum_{k=0}^{\infty} \frac{e^{-mx} \cdot (mx)^k}{k!} \cdot F(k) \right) \\ &= \sum_{k=0}^{\infty} \left(\frac{d}{dx} e^{-mx} x^k \right) \cdot \frac{m^k \cdot F(k)}{k!} \end{aligned}$$

We'll use the linearity of derivatives. To do so, we first compute the derivative of the above coefficients.

$$\frac{d}{dx} e^{-mx} x^k = \begin{cases} -me^{-mx} \cdot x^k & k = 0 \\ -me^{-mx} \cdot x^k + ke^{-mx} x^{k-1} & k \geq 1 \end{cases}$$

Therefore,

$$\begin{aligned} f'(x) &= -m \cdot \sum_{k=0}^{\infty} \frac{e^{-mx} \cdot (mx)^k}{k!} \cdot F(k) + \sum_{k=1}^{\infty} \frac{ke^{-mx} (mx)^{k-1} \cdot m}{k!} \cdot F(k) \\ &= m \cdot \left(-\sum_{k=0}^{\infty} \frac{e^{-mx} \cdot (mx)^k}{k!} \cdot F(k) + \sum_{k=0}^{\infty} \frac{e^{-mx} (mx)^k \cdot m}{k!} \cdot F(k+1) \right) \\ &= m \cdot \left(\sum_{k=0}^{\infty} \frac{e^{-mx} \cdot (mx)^k}{k!} \cdot (F(k+1) - F(k)) \right) \\ &= \mathbb{E}_{\mathbf{k} \sim \text{Poi}(mx)} [\Delta F(\mathbf{k})], \end{aligned}$$

as desired. □

As a consequence, we give the main theorem of this subsection, a formal version of [Theorem 2](#)

Theorem 7 (Formal version of [Theorem 2](#)). *For any $m \in \mathbb{N}$ and convex $F : \mathbb{Z}_{\geq 0} \rightarrow [0, 1]$ satisfying, for any $k_1, k_2 \in \mathbb{Z}_{\geq 0}$,*

$$\Delta^{(2)} F(k_1) \cdot F(k_2) + F(k_1) \cdot \Delta^{(2)} F(k_2) \leq 2 \cdot \Delta F(k_1) \cdot \Delta F(k_2) \quad (20)$$

$$\Delta^{(3)} F(k_1) \cdot \Delta F(k_2) + \Delta F(k_1) \cdot \Delta^{(3)} F(k_2) \leq 2 \cdot \Delta^{(2)} F(k_1) \cdot \Delta^{(2)} F(k_2) \quad (21)$$

Let

$$\Gamma := 1 - \sum_{n=0}^{\infty} \frac{m^n}{(m+1)^{n+1}} \cdot F(n), \quad (22)$$

and suppose that

$$m \leq \frac{-m \cdot \Delta F(0) - \Gamma}{\Gamma + (1 - \Gamma) \cdot m \cdot \Delta F(0)}. \quad (23)$$

Then, if there an (F, m) -discrete OCS, there is an Γ -competitive algorithm for edge-weighted online bipartite matching.

Proof. Apply [Lemma 8.1](#) to construct an (f, m) -continuous OCS for f defined [Equation \(19\)](#). We wish to apply [Theorem 6](#) to that continuous OCS. If F is convex, then f is convex by [Lemma 8.5](#). We verify that f and f' are log-concave:

$$\begin{aligned} f''(x) \cdot f(x) - (f'(x))^2 &= \mathbb{E}_{\mathbf{k}_1, \mathbf{k}_2 \sim \text{Poi}(mx)} \left[\Delta^{(2)} F(k_1) \cdot F(k_2) - \Delta F(k_1) \cdot \Delta F(k_2) \right] \leq 0. \\ f'''(x) \cdot f'(x) - (f''(x))^2 &= \mathbb{E}_{\mathbf{k}_1, \mathbf{k}_2 \sim \text{Poi}(mx)} \left[\Delta^{(3)} F(k_1) \cdot \Delta F(k_2) - \Delta^{(2)} F(k_1) \cdot \Delta^{(2)} F(k_2) \right] \leq 0. \end{aligned}$$

Therefore, the desired result holds with

$$\Gamma = 1 - \int_0^\infty e^{-t} \mathbb{E}_{\mathbf{k} \sim \text{Poi}(mt)} [F(\mathbf{k})] dt$$

We just need to show the above simplifies to [Equation \(22\)](#).

$$\begin{aligned} 1 - \Gamma &= \int_0^\infty e^{-t} \sum_{k=0}^\infty \Pr[\text{Poi}(m \cdot t) = k] \cdot F(k) dt \\ &= \int_0^\infty e^{-t} \sum_{k=0}^\infty \frac{e^{-mt} \cdot (mt)^k}{k!} \cdot F(k) dt \\ &= \sum_{k=0}^\infty \frac{m^k}{k!} F(k) \cdot \int_0^\infty e^{-(m+1)t} \cdot t^k dt \\ &= \sum_{k=0}^\infty \frac{m^k}{k!} F(k) \cdot \left(\frac{k!}{(m+1)^{k+1}} \right) \\ &= \sum_{k=0}^\infty \frac{m^k}{(m+1)^{k+1}} \cdot F(k). \end{aligned}$$

Hence, [Theorem 6](#) guarantees the same competitive ratio as in [Theorem 7](#). \square

Remark 3. [Equation \(23\)](#) constrains how large m can be (we should think of $m \cdot \Delta F(0)$ as fixed; it is -1 in our constructions). For that reason, we are forced to set $m \leq 6$ in order to achieve a competitive ratio of more than 0.533. Depending on the construction, the constraint in [Equation \(23\)](#) may not be fully necessary. That constraint comes from the fact that in [Lemma 8.1](#), the continuous OCS is (f, r) -continuous for $r = m$. There may be tighter bounds for r that allow m to be set significantly higher. For our constructions, we believe even in the case of $m \rightarrow \infty$, that r approaches some constant. Since [Equation \(22\)](#) is often increasing in m (it is for various constructions we tried), more tight analysis of r coupled with larger m could lead to an improved competitive ratio.

8.1 Proof of [Theorems 3 and 5](#)

We first prove the following more general result, and then use it for both [Theorems 3 and 5](#).

Lemma 8.6. *For any $m, k^* \in \mathbb{N}$, $c \in [0, 1]$, and $F : \mathbb{Z}_{\geq 0} \rightarrow [0, 1]$ where, for all $k \geq k^*$,*

$$F(k) = F(k^*) \cdot (1 - c)^{k - k^*}. \tag{24}$$

Suppose that [Equations \(20\)](#) and [\(21\)](#) hold for all $k_1, k_2 < k^*$, that for all $k < k^*$, that $\Delta^2 F(k) \geq 0$ and

$$\Delta^{(2)} F(k) + 2c \cdot \Delta F(k) + c^2 \cdot F(k) \leq 0 \quad (25)$$

$$\Delta^{(3)} F(k) + 2c \cdot \Delta^{(2)} F(k) + c^2 \cdot \Delta F(k) \geq 0. \quad (26)$$

Let,

$$\Gamma := 1 - \left(\sum_{n=0}^{k^*-1} \frac{m^n}{(m+1)^{n+1}} \cdot F(n) \right) - \left(\frac{m}{m+1} \right)^{k^*} \cdot \frac{1}{1+m \cdot c} \cdot F(k^*), \quad (27)$$

and suppose that

$$m \leq \frac{-m \cdot \Delta F(0) - \Gamma}{\Gamma + (1 - \Gamma) \cdot m \cdot \Delta F(0)}. \quad (28)$$

Then, if there is an (F, m) -discrete OCS, there is an Γ -competitive algorithm for edge-weighted online bipartite matching.

Proof. To verify the conditions of [Theorem 7](#), we need to compute the derivatives of F in the case where $k \geq k^*$.

$$\Delta^{(\ell)} F(k) = (-c)^\ell \cdot F(k) \quad \text{for } k \geq k^*$$

Therefore, the condition $\Delta^{(2)} F(k) \geq 0$ for $k < k^*$ is enough to guarantee that F is convex.

We need to show [Equations \(20\)](#) and [\(21\)](#) holds in the case where exactly one of k_1 and k_2 is at least k^* and in the case where both are at least k^* . In the case where $k_1, k_2 \geq k^*$, [Equation \(20\)](#) simplifies to

$$c^2 \cdot F(k_1) \cdot F(k_2) + c^2 \cdot F(k_1) \cdot F(k_2) \leq 2c^2 \cdot F(k_1) \cdot F(k_2)$$

and [Equation \(21\)](#) to

$$c^4 \cdot F(k_1) \cdot F(k_2) + c^4 \cdot F(k_1) \cdot F(k_2) \leq 2c^4 \cdot F(k_1) \cdot F(k_2)$$

both of which hold with equality. Consider the other case, where exactly one of k_1, k_2 is less than k^* . Without loss of generality, let it be k_2 . Then, [Equation \(20\)](#) requires

$$c^2 F(k_1) F(k_2) + F(k_1) \Delta^{(2)} F(k_2) \leq 2(-c) F(k_1) \Delta F(k_2)$$

which is equivalent to [Equation \(25\)](#). Similarly [Equation \(21\)](#) requires

$$-c^3 \cdot F(k_1) \cdot \Delta F(k_2) - c \cdot F(k_1) \cdot \Delta^{(3)} F(k_2) \leq 2c^2 \cdot F(k_1) \cdot \Delta^{(2)} F(k_2)$$

which is equivalent to Equation (26). Therefore, we can apply Theorem 7 to determine the competitive ratio. First, we compute

$$\begin{aligned}
\sum_{n=k^*}^{\infty} \frac{m^n}{(m+1)^{n+1}} \cdot F(n) &= \sum_{n=k^*}^{\infty} \frac{m^n}{(m+1)^{n+1}} \cdot F(k^*) \cdot (1-c)^{n-k^*} \\
&= \frac{\frac{m^{k^*}}{(m+1)^{k^*+1}} \cdot F(k^*)}{1 - \frac{m}{m+1} \cdot (1-c)} \\
&= \left(\frac{m}{m+1}\right)^{k^*} \cdot \frac{1}{(m+1) - m \cdot (1-c)} \cdot F(k^*) \\
&= \left(\frac{m}{m+1}\right)^{k^*} \cdot \frac{1}{1 + m \cdot c} \cdot F(k^*)
\end{aligned}$$

By Equation (22), the competitive ratio is,

$$\begin{aligned}
\Gamma &= 1 - \sum_{n=0}^{\infty} \frac{m^n}{(m+1)^{n+1}} \cdot F(n) \\
&= 1 - \left(\sum_{n=0}^{k^*-1} \frac{m^n}{(m+1)^{n+1}}\right) - \left(\frac{m}{m+1}\right)^{k^*} \cdot \frac{1}{1 + m \cdot c} \cdot F(k^*).
\end{aligned}$$

□

Next we prove the following theorem, restated for convenience.

Theorem 3. *For any $\gamma \in (0, 1)$, if there is a γ -OCS, there is an algorithm for edge-weighted online bipartite matching with a competitive ratio of*

$$\Gamma = \frac{3 + 2\gamma}{6 + 3\gamma}. \quad (1)$$

Proof. Recall that a γ -OCS is $(F, 2)$ -discrete OCS for F ,

$$F(k) := 2^{-k} \cdot (1 - \gamma)^{(k-1)_+}.$$

F has the form in Equation (24) for $k^* = 1$ and $c = \frac{1+\gamma}{2}$. We compute discrete derivatives for $k = 0$:

$$\begin{aligned}
F(0) &= 1 \\
\Delta F(0) &= -\frac{1}{2} \\
\Delta^{(2)} F(0) &= \frac{1 - \gamma}{4} \\
\Delta^{(3)} F(0) &= \frac{\gamma^2 + 4\gamma - 1}{8}
\end{aligned}$$

The second derivative is positive for all $\gamma \in (0, 1)$, so F is convex. We verify Equations (20) and (21) for all $k_1, k_2 < k^*$. The only such case is when $k_1 = k_2 = 0$ in which case we need the following

n	0	1	2	3	4	5	6	7	8	9	$n \geq 10$
$F(n)$	1.0	0.833	0.677	0.54	0.426	0.333	0.260	0.201	0.156	0.121	$0.093 \cdot 0.773^{n-10}$

Figure 5: We prove the existence of an $(F, 6)$ -discrete OCS for the F described in this table. The construction is given in [Sections 3](#) and [10](#) with the hyperparameter $p = 0.48$.

quantities to be nonpositive.

$$2 \cdot \frac{1-\gamma}{4} - 2 \left(-\frac{1}{2}\right)^2 = -\frac{\gamma}{2} \leq 0.$$

$$2 \cdot \left(-\frac{1}{2}\right) \cdot \frac{\gamma^2 + 4\gamma - 1}{8} - 2 \cdot \left(\frac{1-\gamma}{4}\right)^2 = -\frac{\gamma(\gamma+1)}{4} \leq 0.$$

We verify [Equations \(25\)](#) and [\(26\)](#) for $k = 0$

$$\frac{1-\gamma}{4} + 2 \left(\frac{1+\gamma}{2}\right) \left(-\frac{1}{2}\right) + \left(\frac{1+\gamma}{2}\right)^2 \cdot 1 = \frac{\gamma \cdot (\gamma-1)}{4} \leq 0$$

$$\frac{\gamma^2 + 4\gamma - 1}{8} + 2 \left(\frac{1+\gamma}{2}\right) \cdot \frac{1-\gamma}{4} + \left(\frac{1+\gamma}{2}\right)^2 \cdot \left(-\frac{1}{2}\right) = \frac{\gamma(1-\gamma)}{4} \geq 0$$

Then, we compute the competitive ratio.

$$\begin{aligned} \Gamma &= 1 - \left(\sum_{n=0}^{k^*-1} \frac{m^n}{(m+1)^{n+1}} \cdot F(n) \right) - \left(\frac{m}{m+1} \right)^{k^*} \cdot \frac{1}{1+m \cdot c} \cdot F(k^*) \\ &= 1 - \frac{1}{3} \cdot 1 - \frac{2}{3} \cdot \frac{1}{2+\gamma} \cdot \frac{1}{2} \\ &= \frac{3+2\gamma}{6+3\gamma} \end{aligned}$$

Lastly, we need to verify [Equation \(28\)](#). Since $\Delta F(0) = -\frac{1}{m}$, the upper bound on m simplifies to

$$\frac{1-\Gamma}{\Gamma-(1-\Gamma)} = \frac{1-\Gamma}{2\Gamma-1}$$

This is a decreasing function of Γ . The largest possible value for Γ is $\frac{5}{9}$, corresponding to when $\gamma = 1$. In the case, the above simplifies to 4. Since $m = 2 \leq 4$, [Equation \(28\)](#) holds. \square

[Theorem 5](#) is a consequence of [Lemma 8.6](#) and the existence of an $(F, 6)$ -discrete OCS for the $F : \mathbb{Z}_{\geq 0} \rightarrow [0, 1]$ described in [Figure 5](#). In a publicly available Colab notebook¹, we verify that OCS has the properties required by [Lemma 8.6](#) and that the competitive ratio computed by [Equation \(27\)](#) is $\Gamma \geq 0.5368$.

¹To see our code, go to <https://colab.research.google.com/drive/1yQErphKVkwPPXsUWGT2b-nIPPaLBtNh?usp=sharing>

9 From win distributions to discrete OCSes

In this section, we prove the following meta theorem, restated for convenience. Recall the pseudocode for $\text{OCS}_{\mathcal{W},\mathcal{T}}$ is given in [Figure 2](#).

Theorem 4 (Quality of our meta algorithm). *Let \mathcal{W} be a win-distribution with the following three properties.*

1. **Negatively associated:** *For any $n \geq 1$ and $\mathbf{x} \sim \mathcal{W}$, the variables $\mathbf{x}_1, \dots, \mathbf{x}_n$ are negative associated.*
2. **Shift-invariant:** *For any $n \geq 1$ and $\mathbf{x} \sim \mathcal{W}$, the distribution of $(\mathbf{x}_1, \mathbf{x}_2, \dots)$ is identical to that of $(\mathbf{x}_n, \mathbf{x}_{n+1}, \dots)$.*
3. **Sufficiently small:** *For each $r \in [m-1]$ and $t \in [0, 1]$, defining $\mathbf{w}^{(r)}$ to be the random variable representing the distribution of the desired win probability for a vertex of multiplicity r ,*

$$\mathbf{w}^{(r)} \sim \left(1 - \prod_{\ell=1}^r (1 - \mathbf{x}_\ell) \right) \quad \text{where} \quad \mathbf{x} \sim \mathcal{W},$$

the following holds

$$\mathbb{E}_{\mathbf{x} \sim \mathcal{W}} \left[\left(\mathbf{w}^{(r)} - t \right)_+ \right] \leq 1 - t - \frac{m-r}{m} \cdot \left(1 - t^{\frac{m}{m-r}} \right) \quad (3)$$

where $(z)_+$ is short hand for $\max(z, 0)$.

Then, there exists some tournament subroutine \mathcal{T} for which $\text{OCS}_{\mathcal{W},\mathcal{T}}$ is a (f, m) -discrete OCS for

$$F(n) := \mathbb{E}_{\mathbf{x} \sim \mathcal{W}} \left[\prod_{\ell=1}^n (1 - \mathbf{x}_\ell) \right]. \quad (4)$$

We break the proof of [Theorem 4](#) into two steps: In [Lemma 9.1](#) we assuming the existence of a tournament subroutine with some desirable properties. Then, in [Lemma 9.12](#), we prove the existence of such a tournament subroutine as long as the win distribution is sufficiently small.

9.1 Proof of [Section 9](#) assuming a good tournament subroutine

We prove the following Lemma.

Lemma 9.1. *Let \mathcal{W} be a win-distribution with the following properties.*

1. **Negatively associated:** *For any $n \geq 1$ and $\mathbf{x} \sim \mathcal{W}$, the variables $\mathbf{x}_1, \dots, \mathbf{x}_n$ are negative associated.*
2. **Shift-invariant:** *For any $n \geq 1$ and $\mathbf{x} \sim \mathcal{W}$, the distribution of $\mathbf{x}_1, \mathbf{x}_2, \dots$ is identical to that of $\mathbf{x}_n, \mathbf{x}_{n+1}, \dots$.*

Let \mathcal{T} be a tournament subroutine with the following properties.

1. **Consistent:** For each $w \in [0, 1]$, the probability i wins round j given that $\mathbf{w}_j(i) = w$ is at least w .
2. **Stochastic ordering:** Whether i wins round j is a stochastically increasing function of $\mathbf{w}_j(i)$ and a stochastically decreasing function of $\mathbf{w}_j(i')$ for each $i' \neq i$.

Then, $\text{OCS}_{\mathcal{W}, \mathcal{T}}$ is a (F, m) -discrete OCS for

$$F(n) := \mathbb{E}_{\mathbf{x} \sim \mathcal{D}} \left[\prod_{\ell=1}^n (1 - \mathbf{x}_\ell) \right].$$

We'll need a variety of desirable properties satisfied by negatively associated variables. All of the below facts have short proofs (see [JDP83]).

Fact 9.2. *If $\mathbf{x}_1, \dots, \mathbf{x}_n$ are independent, then they are negatively associated.*

Fact 9.3. *The union of independent sets of negatively associated random variables is also negatively associated.*

Fact 9.4. *Any subset of a set of negatively associated variables is also negatively associated.*

Fact 9.5. *Increasing functions defined on disjoint subsets of negatively associated random variables are also negatively associated.*

Next, we prove an analogue of **Fact 9.5** for stochastically increasing functions.

Proposition 9.6. *Stochastically increasing functions defined on disjoint subsets of negatively associated random variables are also negatively associated.*

Proof. Let $\mathbf{x} = \mathbf{x}_1, \dots, \mathbf{x}_n$ be negatively associated, $A_1, \dots, A_m \subseteq [n]$ disjoint subsets, and $\mathbf{f}_1, \dots, \mathbf{f}_m$ stochastically increasing functions. We draw $\mathbf{y}_1, \dots, \mathbf{y}_m \stackrel{iid}{\sim} \text{Uniform}(0, 1)$. By **Fact 9.2** and **Fact 9.3**, the set $(\mathbf{x}_1, \dots, \mathbf{x}_n, \mathbf{y}_1, \dots, \mathbf{y}_m)$ is together negatively associated. Then, we define \mathbf{z}_i for each $i \in M$ as,

$$\mathbf{z}_i = \min v \text{ such that } \Pr[\mathbf{f}_i(\mathbf{x}_{A_i}) \leq v] \geq \mathbf{y}_i$$

Then the distribution of \mathbf{z}_i is equivalent to that of $\mathbf{f}_i(\mathbf{x}_{A_i})$. Furthermore, since \mathbf{f}_i is stochastically increasing, \mathbf{z}_i is an increasing function of \mathbf{x}_{A_i} and \mathbf{y}_i . By **Fact 9.5**, $\mathbf{z}_1, \dots, \mathbf{z}_m$ are negatively associated. \square

As a quick corollary of the above,

Corollary 9.7. *Let $\mathbf{x} = \mathbf{x}_1, \dots, \mathbf{x}_n$ be negatively associated. For any stochastically increasing functions $\mathbf{f}_1, \dots, \mathbf{f}_m$ defined on disjoint coordinates of \mathbf{x} , A_1, \dots, A_m respectively,*

$$\mathbb{E} \left[\prod_{i=1}^m \mathbf{f}_i(\mathbf{x}_{A_i}) \right] \leq \prod_{i=1}^m \mathbb{E}[\mathbf{f}_i(\mathbf{x}_{A_i})]$$

Proof. Let \mathbf{z}_i be the output of $\mathbf{f}_i(\mathbf{x}_{A_i})$ for $i = 1, \dots, m$. Then, by **Proposition 9.6**, $\mathbf{z}_1, \dots, \mathbf{z}_m$ are negatively associated. By repeated application of **Definition 6**, we have the desired result. \square

We'll also need that negative association is preserved under negation.²

Proposition 9.8. *Let $\mathbf{x}_1, \dots, \mathbf{x}_n$ be negatively associated. Then $-\mathbf{x}_1, \dots, -\mathbf{x}_n$ are also negatively associated*

Proof. Let f_1, f_2 be increasing functions over disjoint coordinates $A_1, A_2 \subseteq [n]$. We wish to prove

$$\text{Cov}[f_1((-\mathbf{x})_{A_1}), f_2((-\mathbf{x})_{A_2})] \leq 0.$$

Since $-f_1((-\mathbf{x})_{A_1})$ and $-f_2((-\mathbf{x})_{A_2})$ are increasing and disjoint functions of \mathbf{x} and \mathbf{x} is negatively associated,

$$\text{Cov}[-f_1((-\mathbf{x})_{A_1}), -f_2((-\mathbf{x})_{A_2})] \leq 0.$$

The Lemma follows from the fact that $\text{Cov}[-\mathbf{a}, -\mathbf{b}] = \text{Cov}[\mathbf{a}, \mathbf{b}]$ for any random variables \mathbf{a}, \mathbf{b} . \square

With the above properties established, we are able to make progress towards proving [Lemma 9.1](#). First, we establish that the desired win probabilities are negatively associated.

Proposition 9.9. *Under the conditions of [Lemma 9.1](#), for each $i \in L$, the variables $\mathbf{w}_1(i), \mathbf{w}_2(i), \dots$ in [Figure 2](#) are negatively associated.*

Proof. First, we claim that for any $r \geq 1$, the function $1 - \prod_{i=1}^r (1 - x_i)$ is an increasing function of x_1, \dots, x_r . This is by a series of compositions.

$$\begin{aligned} (x_i \mapsto 1 - x_i) & \qquad \qquad \qquad \text{is decreasing} \\ \implies ((x_1, \dots, x_r) \mapsto \prod_{i=1}^r (1 - x_i)) & \qquad \qquad \qquad \text{is decreasing} \\ \implies ((x_1, \dots, x_r) \mapsto 1 - \prod_{i=1}^r (1 - x_i)) & \qquad \qquad \qquad \text{is increasing} \end{aligned}$$

Therefore, $\mathbf{w}_j(i)$ are increasing functions of disjoint subsets of $(\mathbf{x}_1(i), \mathbf{x}_2(i), \dots)$. The desired result follows from [Fact 9.5](#). \square

Lemma 9.10. *Under the conditions of [Lemma 9.1](#), for any $i \in L$ and disjoint subsets of time steps $S_1, \dots, S_k \subseteq [T]$*

$$\Pr[i \text{ never a winner in } \bigcup_{\ell=1}^k S_\ell] \leq \prod_{\ell=1}^k \Pr[i \text{ never a winner in } S_\ell]$$

Proof. According to the properties of \mathcal{T} , whether i never wins in S_ℓ is a stochastically increasing function of $(-\mathbf{w}_j(i))$ and of $\mathbf{w}_j(i')$ for each $i' \neq i, j \in S_\ell$. By [Proposition 9.8](#) and [Proposition 9.9](#), the variables $(-\mathbf{w}_1(i), -\mathbf{w}_2(i), \dots)$ are negatively associated. Applying [Fact 9.3](#), we have that

$$\bigcup_{j=1}^{\infty} \left(\{-\mathbf{w}_j(i)\} \cup \bigcup_{i' \neq i} \mathbf{w}_j(i') \right) \text{ are collectively negatively associated.}$$

The desired results follows [Corollary 9.7](#). \square

²Negative association is only preserved under negation of *all* the variables. Negating only some variables can create a positive correlation.

Lemma 9.10 allows us to prove **Lemma 9.1** by just considering how the OCS behaves on consecutive subsequences rather than on unions of consecutive subsequences. We prove a final ingredient.

Lemma 9.11. *Under the conditions of **Lemma 9.1**, for any $i \in L$, subset of time steps $S \subseteq [T]$ in which i appears, and desired win probabilities $w_j(i)$ for each $j \in S$,*

$$\Pr[i \text{ never a winner in } S \mid \mathbf{w}_j(i) = w_j(i) \text{ for each } j \in S] \leq \prod_{j \in S} (1 - w_j(i)).$$

Proof. Just as in **Lemma 9.10**,

$$\bigcup_{j=1}^{\infty} \left(\bigcup_{i' \neq i} \mathbf{w}_j(i') \right) \text{ are all negatively associated.}$$

Whether i loses in round j is a stochastically increasing function of $w_j(i')$ for each $i \neq i'$. Applying **Corollary 9.7**,

$$\begin{aligned} \Pr[i \text{ never a winner in } S \mid \mathbf{w}_j(i) = w_j(i) \text{ for each } j \in S] \\ \leq \prod_{j \in S} (\Pr[i \text{ loses in round } j \mid \mathbf{w}_j(i) = w_j(i)]) \end{aligned}$$

By the consistency of \mathcal{T} , $\Pr[i \text{ loses in round } j \mid \mathbf{w}_j(i) = w_j(i)]$ is at most $1 - w_j(i)$, proving the desired result. \square

Finally, we prove **Lemma 9.1**.

*Proof of **Lemma 9.1**.* Our goal is to prove that for any vertex $i \in L$ and disjoint consecutive time steps S_1, \dots, S_k , that

$$\Pr[i \text{ never a winner in } S_1, \dots, S_k] \leq \prod_{\ell=1}^k F(\text{count}_i(S_\ell))$$

where f is as defined in **Equation (4)**. By **Lemma 9.10** and the assumption that \mathcal{W} is shift-invariant, it is enough to prove, for $r_j(i) :=$ the number of appearances of i in A_j ,

$$\Pr[i \text{ never a winner in } S := \{1, \dots, m\}] \leq F \left(\sum_{j=1}^m r_j(i) \right).$$

We compute,

$$\begin{aligned}
\Pr[i \text{ never a winner in } S] &= \mathbb{E}_{\mathbf{w}_1(i), \dots, \mathbf{w}_m(i)} [\Pr[i \text{ never a winner in } S \mid \mathbf{w}_1(i), \dots, \mathbf{w}_m(i)]] \\
&\leq \mathbb{E}_{\mathbf{w}_1(i), \dots, \mathbf{w}_m(i)} \left[\prod_{j \in S} (1 - \mathbf{w}_j(i)) \right] && \text{(Lemma 9.11)} \\
&= \mathbb{E}_{\mathbf{x}^{(i)} \sim \mathcal{W}} \left[\prod_{j \in S} \left(\prod_{\ell=0}^{r_j(i)-1} (1 - \mathbf{x}_{k_j(i)+\ell}) \right) \right] && \text{(Equation (2))} \\
&= \mathbb{E}_{\mathbf{x}^{(i)} \sim \mathcal{W}} \left[\prod_{\ell=1}^{r_1(i)+\dots+r_m(i)} (1 - \mathbf{x}_\ell) \right] && \text{(Combine products)} \\
&= F \left(\sum_{j \in S} r_j(i) \right) && \text{(Equation (4))}
\end{aligned}$$

as desired. \square

9.2 Design of the tournament subroutine

In this subsection, we prove the following Lemma which completes the proof of [Theorem 4](#).

Lemma 9.12 (Sufficiently small \implies good \mathcal{T}). *Let \mathcal{W} be a win-distribution for which the following holds. For each $r \in [m-1]$ define $\mathbf{w}^{(r)}$ to be the random variable representing the distribution of the desired win probability of a vertex of multiplicity r ,*

$$\mathbf{w}^{(r)} \sim \left(1 - \prod_{\ell=1}^r (1 - \mathbf{x}_\ell) \right) \quad \text{where} \quad \mathbf{x} \sim \mathcal{W},$$

If, for each $t \in [0, 1]$ and $r \in [m-1]$,

$$\mathbb{E}_{\mathbf{x} \sim \mathcal{W}} \left[\left(\mathbf{w}^{(r)} - t \right)_+ \right] \leq 1 - t - \frac{m-r}{m} \cdot \left(1 - t^{\frac{m}{m-r}} \right), \quad (29)$$

then there is a tournament subroutine, \mathcal{T} with the following properties.

1. **Consistent:** For each $w \in [0, 1]$, the probability i wins round j given that $\mathbf{w}_j(i) = w$ is at least w .
2. **Stochastic ordering:** The probability i wins a round is a stochastically increasing function of $\mathbf{w}_j(i)$ and a stochastically decreasing function of $\mathbf{w}_j(i')$ for each $i' \neq i$.

Our tournament subroutine will be simple: We'll design stochastically increasing strength functions $\mathcal{S}_1, \dots, \mathcal{S}_{m-1}$ that map a desired win probability to a "strength." For each $i \in A_j$, we set its strength to $\mathcal{S}_{r_j(i)}(\mathbf{w}_j(i))$, and the vertex with the highest strength wins the round. We give pseudocode for this subroutine in [Figure 6](#).

As long as the strength functions are stochastically increasing, the stochastic ordering criteria of [Lemma 9.12](#) will hold. On the other hand, consistency is more challenging. The following fundamental question underlies this subsection.

$\mathcal{T}_{\mathcal{S}_1, \dots, \mathcal{S}_{m-1}}$

1. If there is a vertex with $r_j(i) = m$, then select i as the winner.
2. Otherwise, for each $i \in A_j$, sample a strength,

$$\mathbf{s}_j(i) := \mathcal{S}_{r_j(i)}(\mathbf{w}_j(i)),$$

and choose the i maximizing $\mathbf{s}_j(i)$ as the winner.

Figure 6: A tournament subroutine for returning a winner given desired win probabilities $\mathbf{s}_j(i)$ and multiplicities $r_j(i)$. This subroutine is parameterized by strength functions $\mathcal{S}_1, \dots, \mathcal{S}_{m-1}$.

For which distributions of $\mathbf{w}^{(r)}$ is it possible to design stochastically increasing strength functions such that the probability i wins round j given $\mathbf{w}_j(i) = w$ is at least w . (\diamond)

Consider the case where $r = 1$. Since there can only be one winner in a round, for (\diamond) to hold, it must be the case that $\mathbb{E}[\mathbf{w}^{(1)}] \leq \frac{1}{m}$. It turns out, that condition is not sufficient. Suppose that $\mathbf{w}^{(1)}$ is 1 with probability $\frac{1}{m}$ and 0 otherwise. Then, there is a nonzero probability that 2 (or more) vertices will both declare that they should win with probability 1, making (\diamond) impossible. This example shows that not only do we want the expectation of $\mathbf{w}^{(1)}$ to be small, we want its distribution to not be too spread out.

We will show that these two notions – how large and spread out $\mathbf{w}^{(r)}$ is – fully capture when (\diamond) is possible. Recall the following definition which formalizes the notion of “large and spread out.”

Definition 10 (Increasing convex order). *Let \mathbf{a}, \mathbf{b} be two univariate random variables such that,*

$$\mathbb{E}[\phi(\mathbf{a})] \leq \mathbb{E}[\phi(\mathbf{b})]$$

for all increasing convex functions $\phi: \mathbb{R} \rightarrow \mathbb{R}$ for which the above expectations exist. Then, we say \mathbf{a} is smaller than \mathbf{b} in the increasing convex order and denote this,

$$\mathbf{a} \preceq_{\text{icx}} \mathbf{b}.$$

The following Lemma formalizes our intuition that as long as the desired win probabilities are sufficiently small and not spread out, (\diamond) holds.

Lemma 9.13 (Increasing convex order \rightarrow (\diamond)). *For any random variables $\mathbf{w}^{(1)}, \dots, \mathbf{w}^{(m-1)}$ and $\mathbf{w}'^{(1)}, \dots, \mathbf{w}'^{(m-1)}$, if*

$$\mathbf{w}^{(r)} \preceq_{\text{icx}} \mathbf{w}'^{(r)} \quad \text{for each } r = 1, \dots, m-1,$$

and (\diamond) holds for $(\mathbf{w}'^{(1)}, \dots, \mathbf{w}'^{(m-1)})$, then it also holds for $(\mathbf{w}^{(1)}, \dots, \mathbf{w}^{(m-1)})$.

Our proof of [Lemma 9.13](#) leverages a powerful theorem of Müller and Rüschemdorf.

Theorem 8 ([MR01]). *For univariate random variables \mathbf{a}, \mathbf{b} , the following are equivalent.*

1. \mathbf{a} is smaller than \mathbf{b} in the increasing convex order.
2. There is a coupling of $(\hat{\mathbf{a}}, \hat{\mathbf{b}})$ such that the marginal distribution of $\hat{\mathbf{a}}$ and $\hat{\mathbf{b}}$ are the same as those of \mathbf{a} and \mathbf{b} respectively, and, for any choice of a ,

$$\mathbb{E}[\hat{\mathbf{b}} \mid \hat{\mathbf{a}}] \geq \hat{\mathbf{a}}.$$

Furthermore, the stochastic function mapping $a \mapsto (\hat{\mathbf{b}} \mid \hat{\mathbf{a}} = a)$ is stochastically increasing.

Theorem 8 is a generalization of a classic result by Strassen [Str65] who proved the same result without the restriction that $a \mapsto (\hat{\mathbf{b}} \mid \hat{\mathbf{a}} = a)$ be stochastically increasing. We'll need the stronger result since we want our strength functions to be stochastically increasing.

Proof of Lemma 9.13. For each $r \in [m-1]$, let $(\hat{\mathbf{w}}^{(r)}, \hat{\mathbf{w}}'^{(r)})$ be the coupling guaranteed to exist by **Theorem 8** and let $\mathcal{S}'_1, \dots, \mathcal{S}'_{m-1}$ be the strength functions satisfying (\diamond) for $\mathbf{w}'^{(1)}, \dots, \mathbf{w}'^{(m-1)}$. Define, for each $r \in [m-1]$,

$$\mathcal{S}_r(w) := \mathcal{S}'_r(\hat{\mathbf{w}}') \quad \text{where} \quad \hat{\mathbf{w}}' \sim (\hat{\mathbf{w}}'^{(r)} \mid \hat{\mathbf{w}}^{(r)} = w).$$

We claim that $\mathcal{S}_1, \dots, \mathcal{S}_{m-1}$ are strength functions satisfying (\diamond) for $(\mathbf{w}^{(1)}, \dots, \mathbf{w}^{(m-1)})$. First, since the composition of stochastically increasing functions is stochastically increasing, \mathcal{S}_r is stochastically increasing for all $r \in [m-1]$. Second, the probability i wins a round with strength w is at least $\mathbb{E}[\hat{\mathbf{w}}'^{(r)} \mid \hat{\mathbf{w}}^{(r)} = w]$, which is at least w . \square

Remark 4 (Computational efficiency of **Lemma 9.13**). Computing \mathcal{S}_r requires knowing the coupling $(\hat{\mathbf{w}}^{(r)}, \hat{\mathbf{w}}'^{(r)})$. If $\hat{\mathbf{w}}^{(r)}$ and $\hat{\mathbf{w}}'^{(r)}$ are discrete random variables, finding a desired coupling is simply a linear program. If they are continuous, they can be discretized to any desired accuracy and then treated as discrete.

In order to use **Lemma 9.13**, we need to be able to verify when one variable is smaller, in the increasing convex order, than another. We use the following well known characterization.

Fact 9.14 ([SS07]). *For univariate random variables \mathbf{a}, \mathbf{b} , the following are equivalent.*

1. \mathbf{a} is smaller than \mathbf{b} in the increasing convex order.
2. For every $t \in \mathbb{R}$,

$$\mathbb{E}[(\mathbf{a} - t)_+] \leq \mathbb{E}[(\mathbf{b} - t)_+].$$

We complete the proof of **Lemma 9.12** by giving a particular choice for $(\mathbf{w}^{(1)}, \dots, \mathbf{w}^{(m-1)})$ on which (\diamond) holds.

Lemma 9.15. *For each $r \in [m-1]$, let $\mathbf{w}^{(r)}$ be the random variable supported on $[0, 1]$ satisfying*

$$\mathbb{E}[(\mathbf{w}^{(r)} - t)_+] = 1 - t - \frac{m-r}{m} \cdot \left(1 - t^{\frac{m}{m-r}}\right) \quad \forall t \in [0, 1] \quad (30)$$

Then, (\diamond) holds for $(\mathbf{w}^{(1)}, \dots, \mathbf{w}^{(m-1)})$

Proof. Straightforward computation verifies that Equation (30) is equivalent to each $\mathbf{w}^{(r)}$ having the CDF.

$$\Pr[\mathbf{w}^{(r)} \leq t] = t^{\frac{r}{m-r}} \quad \forall t \in [0, 1].$$

Consider the strength functions, for each $r \in [m - 1]$,

$$\mathcal{S}_r(w) := w^{\frac{1}{m-r}}.$$

For each $r \in [m - 1]$, we have that

$$\Pr[\mathcal{S}_r(\mathbf{w}^{(r)}) \leq t] = \Pr[\mathbf{w}^{(r)} \leq t^{m-r}] = t^r.$$

Therefore, $\mathcal{S}_r(\mathbf{w}^{(r)})$ is distributed according to the max of r independent variables drawn from Uniform(0, 1). A vertex i is selected in a round if its strength is larger than the strengths of each of the other vertices. If i has multiplicity $r(i)$, then the max strength of the other vertices is distributed according to the maximum of $m - r(i)$ independent uniforms. Therefore,

$$\begin{aligned} \Pr[i \text{ selected with strength } w] &= \Pr[\max \text{ of } (m - r(i)) \text{ independent uniforms} < \mathcal{S}_{r(i)}(w)] \\ &= \mathcal{S}_{r(i)}(w)^{m-r(i)} \\ &= w, \end{aligned}$$

Furthermore, \mathcal{S}_r is increasing for all $r \in [m - 1]$. Therefore, both criteria of (\diamond) hold. \square

10 Constructing a win distribution

In this section we will construct a distribution $(\mathbf{x}_1, \mathbf{x}_2, \dots) \sim \mathcal{W}$ that meets the three criteria of Theorem 4 and also leads to a good competitive ratio. We do this in two steps. First, we construct a “seed distribution” \mathcal{D} that returns an infinite sequence $(\mathbf{y}_1, \mathbf{y}_2, \dots)$ of random variables that are negatively associated and shift-invariant, but the values \mathbf{y}_j do not represent probabilities (they are unbounded). Then, for a carefully chosen increasing function W , we set $\mathbf{x}_k = W(\mathbf{y}_k)$ to ensure the third criteria of Theorem 4 holds. We can think of W as mapping an output of \mathcal{D} to a desired win probability.

In the remainder of this section, p is a hyperparameter than impacts the distribution of \mathcal{D} and \mathcal{W} . For concreteness, we will eventually set $p = 0.48$ to optimize the competitive ratio, a value we selected using brute-force search.

Definition 13 (\mathcal{D}). *A sample $(\mathbf{y}_1, \mathbf{y}_2, \dots) \sim \mathcal{D}$ is generated by the following process, where each random choice is independent.*

1. Initialize $\mathbf{z}_1 \sim \text{Geo}(p)$.
2. For $k = 1, 2, \dots$,
 - (a) With probability p , set $\mathbf{y}_k = \mathbf{z}_k$ and then reset z , setting $\mathbf{z}_{k+1} \leftarrow 0$.
 - (b) Otherwise, set $\mathbf{y}_k = -1$ and increment z , setting $\mathbf{z}_{k+1} \leftarrow \mathbf{z}_k + 1$

It will often be convenient for us to refer to the following variants of \mathcal{D} .

Definition 14. We use the following notation to specify variants of \mathcal{D} .

1. For any $v \geq 0$, we use \mathcal{D}_v to refer to the generative process of [Definition 13](#) where \mathbf{z}_1 is initialized to v instead of drawn from $\text{Geo}(p)$.
2. For any $n \geq 1$, we use $\mathcal{D}^{(n)}$ to refer to first n element of \mathcal{D} ; i.e. to $\mathbf{y}_1, \dots, \mathbf{y}_n$ where $\mathbf{y} \sim \mathcal{D}$.
3. We use $\mathcal{D}(y, z)$ to refer to the distribution that outputs two infinite sequences, one corresponding to the $(\mathbf{y}_1, \mathbf{y}_2, \dots)$ and the other to the $(\mathbf{z}_1, \mathbf{z}_2, \dots)$ in [Definition 13](#).

We'll also mix and match the above notation. For example, we might use $\mathbf{y}, \mathbf{z} \sim \mathcal{D}^{(n)}(y, z)$ as shorthand for \mathbf{y} and \mathbf{z} are each n -tuples containing $\mathbf{y}_1, \dots, \mathbf{y}_n$ and $\mathbf{z}_1, \dots, \mathbf{z}_n$ respectively.

By examining the generative process in [Definition 13](#), we observe that \mathcal{D} is *almost* memoryless.

Fact 10.1 (\mathcal{D} is memoryless except for z). For any $k, v \geq 0$ and $\mathbf{y}, \mathbf{z} \sim \mathcal{D}(y, z)$, the following distributions are identical.

1. The distribution of $(\mathbf{y}_k, \mathbf{y}_{k+1}, \dots)$ conditioned on $\mathbf{z}_k = v$.
2. The distribution \mathcal{D}_v .

First, we prove that \mathcal{D} is shift-invariant. This amounts to proving that for any $k \geq 1$, \mathbf{z}_k is distributed according to a $\text{Geo}(p)$.

Proposition 10.2. Let $\mathbf{y}, \mathbf{z} \sim \mathcal{D}(y, z)$. Then, for any $k \geq 1$, \mathbf{z}_k is distributed according to a $\text{Geo}(p)$ distribution,

$$\Pr[\mathbf{z}_k = v] = (1 - p)^v p.$$

Proof. By induction on k ; \mathbf{z}_1 is initialized from a geometric distribution, so [Proposition 10.2](#) holds for $k = 1$.

For the inductive step, consider some $k > 1$, $\mathbf{z}_k = 0$ only if it was reset at that round, which happens with probability p . For $v \geq 1$

$$\begin{aligned} \Pr[\mathbf{z}_k = v] &= (1 - p) \cdot \Pr[\mathbf{z}_{k-1} = v - 1] \\ &= (1 - p) \cdot (1 - p)^{v-1} \cdot p \\ &= (1 - p)^v \cdot p. \end{aligned}$$

Therefore, $\mathbf{z}_k \sim \text{Geo}(p)$. □

As a corollary of [Fact 10.1](#) and [Proposition 10.2](#), the following holds.

Corollary 10.3 (\mathcal{D} is shift-invariant). For any $k \geq 1$ and $\mathbf{y} \sim \mathcal{D}$, the distribution of $\mathbf{y}_1, \mathbf{y}_2, \dots$ is identical to that of $\mathbf{y}_k, \mathbf{y}_{k+1}, \dots$

10.1 \mathcal{D} is negatively associated

We devote this subsection to establishing that \mathcal{D} is negatively associated. Recall the definition for negatively associated variables.

Definition 6 (Negative association [JDP83]). *Random variables $\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n$ are said to be negatively associated if, for every pair of disjoint subsets $A_1, A_2 \subseteq [n]$ and nondecreasing functions f_1, f_2 ,*

$$\text{Cov}[f_1(\mathbf{x}_{A_1}), f_2(\mathbf{x}_{A_2})] \leq 0$$

where \mathbf{x}_A is the vector $(\mathbf{x}_{i_1}, \mathbf{x}_{i_2}, \dots, \mathbf{x}_{i_{|A|}})$ when $A = \{i_1, i_2, \dots, i_{|A|}\}$.

For starters, we prove (a strengthening of) the above definition holds for $\mathcal{D}^{(n)}$ and $\mathcal{D}_0^{(n)}$ if we restrict $A_1 = \{n\}$.

Lemma 10.4. *For any $n \geq 1$ and indices $S \subseteq [n-1]$, let $\mathbf{y}_1, \dots, \mathbf{y}_n$ be drawn as*

$$(\mathbf{y}_1, \dots, \mathbf{y}_n) \sim (\mathcal{D}^{(n)} \mid \mathbf{y}_i = -1 \text{ for each } i \in S)$$

Then, for any increasing functions $f_1 : \mathbb{Z} \rightarrow \mathbb{R}, f_2 : \mathbb{Z}^{n-1} \rightarrow \mathbb{R}$,

$$\text{Cov}_{\mathbf{y}_1, \dots, \mathbf{y}_n} [f_1(\mathbf{y}_n), f_2(\mathbf{y}_1, \dots, \mathbf{y}_{n-1})] \leq 0$$

The same is true when using $\mathcal{D}_0^{(n)}$ rather than $\mathcal{D}^{(n)}$.

We only prove [Lemma 10.4](#) for $\mathcal{D}^{(n)}$ as the proof for $\mathcal{D}_0^{(n)}$ is identical. Throughout the following proof, the distribution on $\mathbf{y}_1, \dots, \mathbf{y}_n$ is by default $(\mathcal{D}^{(n)} \mid \mathbf{y}_i = -1 \text{ for each } i \in S)$.

Proof. Without loss of generality, we can assume $f_1(-1) = 0$; otherwise, add an appropriate constant to f_1 so that it's true. Since f_1 is an increasing function of \mathbf{y}_n and $\mathbf{y}_n \geq -1$, we can write,

$$f_1(k) = \sum_{j=0}^k g(j)$$

for an appropriate nonnegative function $g : \mathbb{Z}_{\geq 0} \rightarrow \mathbb{R}_{\geq 0}$. We then expand the first term of the desired covariance.

$$\begin{aligned} \mathbb{E}_{\mathbf{y}_1, \dots, \mathbf{y}_n} [f_1(\mathbf{y}_n) \cdot f_2(\mathbf{y}_1, \dots, \mathbf{y}_{n-1})] &= \sum_{k=-1}^{\infty} f_1(k) \cdot \Pr[\mathbf{y}_n = k] \cdot \mathbb{E} [f_2(\mathbf{y}_1, \dots, \mathbf{y}_{n-1}) \mid \mathbf{y}_n = k] \\ &= \sum_{k=-1}^{\infty} \sum_{j=0}^k g(j) \cdot \Pr[\mathbf{y}_n = k] \cdot \mathbb{E} [f_2(\mathbf{y}_1, \dots, \mathbf{y}_{n-1}) \mid \mathbf{y}_n = k] \\ &= \sum_{j=0}^{\infty} \sum_{k=j}^{\infty} g(j) \cdot \Pr[\mathbf{y}_n = k] \cdot \mathbb{E} [f_2(\mathbf{y}_1, \dots, \mathbf{y}_{n-1}) \mid \mathbf{y}_n = k] \\ &= \sum_{j=0}^{\infty} g(j) \cdot \Pr[\mathbf{y}_n \geq j] \cdot \mathbb{E} [f_2(\mathbf{y}_1, \dots, \mathbf{y}_{n-1}) \mid \mathbf{y}_n \geq j] \end{aligned}$$

We claim that $\mathbb{E} [f_2(\mathbf{y}_1, \dots, \mathbf{y}_{n-1}) \mid \mathbf{y}_n \geq j]$ is a decreasing function of j . First we note that conditioning on $(\mathbf{y}_n \geq j)$ forces that $\mathbf{y}_{n-j} = \dots = \mathbf{y}_{n-j-1} = -1$ and leaves the distribution of $\mathbf{y}_1, \dots, \mathbf{y}_{n-j-2}$ unchanged.

Suppose we draw $\mathbf{y}_1, \dots, \mathbf{y}_{n-1}$ conditioned on $(\mathbf{y}_n \geq j)$. Then, we set $\mathbf{y}_{n-(j+1)}$ to -1 and leave all other $\mathbf{y}_1, \dots, \mathbf{y}_{n-1}$ unchanged. This is equivalent to drawing $\mathbf{y}_1, \dots, \mathbf{y}_{n-1}$ conditioned on $(\mathbf{y}_n \geq j+1)$. Importantly, setting $\mathbf{y}_{n-(j+1)} = -1$ can only decrease the value of $f_2(\mathbf{y}_1, \dots, \mathbf{y}_{n-1})$, because f_2 is increasing. Therefore, we have that,

$$\mathbb{E} [f_2(\mathbf{y}_1, \dots, \mathbf{y}_{n-1}) \mid \mathbf{y}_n \geq j] \quad \text{is decreasing in } j. \quad (31)$$

We can now prove the desired result.

$$\begin{aligned} & \mathbb{E} [f_1(\mathbf{y}_n) \cdot f_2(\mathbf{y}_1 \cdot \dots \cdot \mathbf{y}_{n-1})] \\ &= \sum_{j=0}^{\infty} g(j) \cdot \Pr[\mathbf{y}_n \geq j] \cdot \mathbb{E} [f_2(\mathbf{y}_1, \dots, \mathbf{y}_{n-1}) \mid \mathbf{y}_n \geq j] \\ &\leq \sum_{j=0}^{\infty} g(j) \cdot \Pr[\mathbf{y}_n \geq j] \cdot \mathbb{E} [f_2(\mathbf{y}_1, \dots, \mathbf{y}_{n-1}) \mid \mathbf{y}_n \geq -1] \\ &= \left(\sum_{j=0}^{\infty} g(j) \cdot \Pr[\mathbf{y}_n \geq j] \right) \cdot (\mathbb{E} [f_2(\mathbf{y}_1, \dots, \mathbf{y}_{n-1})]) \\ &= \mathbb{E} [f_1(\mathbf{y}_n)] \cdot \mathbb{E} [f_2(\mathbf{y}_1 \cdot \dots \cdot \mathbf{y}_{n-1})]. \end{aligned}$$

Where the inequality follows from [Equation \(31\)](#). □

Next, we prove that $\mathcal{D}^{(n)}$ is negatively associated.

Lemma 10.5. *For any $n \geq 1$, draw $\mathbf{y}_1, \dots, \mathbf{y}_n \sim \mathcal{D}^{(n)}$. Then, $\mathbf{y}_1, \dots, \mathbf{y}_n$ are negatively associated.*

Proof. By induction on n . For $n = 1$, the desired result holds because a single variable is vacuously negatively associated.

For any $n \geq 2$ fix some disjoint sets $A_1, A_2 \subseteq [n]$ and increasing functions $f_1 : \mathbb{Z}^{|A_1|} \rightarrow \mathbb{R}$, $f_2 : \mathbb{Z}^{|A_2|} \rightarrow \mathbb{R}$. Without loss of generality, we can assume that $n \in A_1$ and $n \notin A_2$. We define an ‘‘averaging’’ of f_1 that does *not* depend upon \mathbf{y}_n .

$$\bar{f}_1(\mathbf{y}_{A_1}) := \mathbb{E}_{\mathbf{y} \sim \mathcal{D}^{(n)}} [f_1(\mathbf{y}_{A_1}) \mid \mathbf{y}_i = \mathbf{y}_i \text{ for each } i \in (A_1 \setminus \{n\})]$$

Importantly, neither \bar{f}_1 nor f_2 depend on \mathbf{y}_n , so we can apply the inductive hypothesis.

$$\mathbb{E}[\bar{f}_1(\mathbf{y}_{A_1}) \cdot f_2(\mathbf{y}_{A_2})] \leq \mathbb{E}[\bar{f}_1(\mathbf{y}_{A_1})] \cdot \mathbb{E}[f_2(\mathbf{y}_{A_2})] \quad (32)$$

Furthermore, \bar{f}_1 and f_1 have the same expectation. Our next goal is to show that

$$\mathbb{E}[f_1(\mathbf{y}_{A_1}) \cdot f_2(\mathbf{y}_{A_2})] \leq \mathbb{E}[\bar{f}_1(\mathbf{y}_{A_1}) \cdot f_2(\mathbf{y}_{A_2})]$$

The above two equations will imply the desired result.

Define the “left signature” of some y_1, \dots, y_n as follows. If $y_i = -1$ for all $i \in A_1 \setminus \{n\}$, then the signature is $\mathcal{L}(y_1, \dots, y_n) = \emptyset$. Otherwise, let i^* be the largest $i \in A_1 \setminus \{n\}$ where $y_i \geq 0$ and the signature is $\mathcal{L}(y_1, \dots, y_n) = (y_1, \dots, y_{i^*})$. We claim for any choice of left signature, ℓ , that

$$\mathbb{E}[f_1(\mathbf{y}_{A_1}) \cdot f_2(\mathbf{y}_{A_2}) \mid \mathcal{L}(\mathbf{y}) = \ell] \leq \mathbb{E}[\bar{f}_1(\mathbf{y}_{A_1}) \cdot f_2(\mathbf{y}_{A_2}) \mid \mathcal{L}(\mathbf{y}) = \ell].$$

First consider the case where $\ell \neq \emptyset$. Conditioning upon signature ℓ is equivalent to:

1. Fixing the values of $\mathbf{y}_1, \dots, \mathbf{y}_{i^*}$ to those in ℓ .
2. Conditioning on $\mathbf{y}_j = -1$ for all $j \in A_1 \setminus \{n\}$ where $j > i^*$.

By [Fact 10.1](#), the distribution of $\mathbf{y}_{i^*+1}, \dots, \mathbf{y}_n$ is the same as $\mathcal{D}_0^{(n-i^*)}$ conditioned on some of the elements being -1 . This is a distribution for which we can apply [Lemma 10.4](#). We also need f_1 to just depend on the last bit, \mathbf{y}_n . Fortunately, conditioning on the signature fixes all bits in A_1 except for \mathbf{y}_n so that property is satisfied. Therefore, by [Lemma 10.4](#).

$$\mathbb{E}[f_1(\mathbf{y}_{A_1}) \cdot f_2(\mathbf{y}_{A_2}) \mid \mathcal{L}(\mathbf{y}) = \ell] \leq \mathbb{E}[f_1(\mathbf{y}_{A_1}) \mid \mathcal{L}(\mathbf{y}) = \ell] \cdot \mathbb{E}[f_2(\mathbf{y}_{A_2}) \mid \mathcal{L}(\mathbf{y}) = \ell] \quad (33)$$

In the other case where $\ell = \emptyset$, conditioning on ℓ is equivalent to drawing $\mathbf{y}_1, \dots, \mathbf{y}_n$ from $(\mathcal{D}^{(n)} \mid y_i = -1 \text{ for each } i \in A_1 \setminus \{n\})$. Once again, after this conditioning, f_1 only depends on \mathbf{y}_n , so we can apply [Lemma 10.4](#) and [Equation \(33\)](#) holds.

Next, we wish to substitute in \bar{f}_1 for f_1 . Fixing $\mathcal{L}(\mathbf{y}) = \ell$ fixes all variables in A_1 except for n . By the definition of \bar{f}_1 , we have that

$$\mathbb{E}[f_1(\mathbf{y}_{A_1}) \mid \mathcal{L}(\mathbf{y}) = \ell] = \bar{f}_1(x) \text{ for any } x \text{ where } \mathcal{L}(x) = \ell. \quad (34)$$

We’ll use two aspects of the above equation:

1. That $\bar{f}_1(x)$ has the same expectation as f_1 when conditioning on the signature.
2. That $\bar{f}_1(x)$ is constant when conditioning on the signature.

, Combining the above with [Equation \(33\)](#), where ℓ is drawn from $\mathcal{L}(\mathbf{y})$ for $\mathbf{y} \sim \mathcal{D}^{(n)}$,

$$\begin{aligned} \mathbb{E}[f_1(\mathbf{y}_{A_1}) \cdot f_2(\mathbf{y}_{A_2})] &= \mathbb{E}_{\ell} [\mathbb{E}[f_1(\mathbf{y}_{A_1}) \cdot f_2(\mathbf{y}_{A_2}) \mid \mathcal{L}(\mathbf{y}) = \ell]] && \text{(Law of total expectation)} \\ &\leq \mathbb{E}_{\ell} [\mathbb{E}[f_1(\mathbf{y}_{A_1}) \mid \mathcal{L}(\mathbf{y}) = \ell] \cdot \mathbb{E}[f_2(\mathbf{y}_{A_2}) \mid \mathcal{L}(\mathbf{y}) = \ell]] && \text{(Equation (33))} \\ &= \mathbb{E}_{\ell} [\mathbb{E}[\bar{f}_1(\mathbf{y}_{A_1}) \mid \mathcal{L}(\mathbf{y}) = \ell] \cdot \mathbb{E}[f_2(\mathbf{y}_{A_2}) \mid \mathcal{L}(\mathbf{y}) = \ell]] && \text{(Equation (34))} \\ &= \mathbb{E}_{\ell} [\mathbb{E}[\bar{f}_1(\mathbf{y}_{A_1}) \cdot f_2(\mathbf{y}_{A_2}) \mid \mathcal{L}(\mathbf{y}) = \ell]] && (\bar{f}_1 \text{ constant conditioned on } \ell) \\ &= \mathbb{E}[\bar{f}_1(\mathbf{y}_{A_1}) \cdot f_2(\mathbf{y}_{A_2})] && \text{(Law of total expectation)} \\ &\leq \mathbb{E}[\bar{f}_1(\mathbf{y}_{A_1})] \cdot \mathbb{E}[f_2(\mathbf{y}_{A_2})] && \text{(Equation (32))} \\ &= \mathbb{E}[f_1(\mathbf{y}_{A_1})] \cdot \mathbb{E}[f_2(\mathbf{y}_{A_2})] && (\mathbb{E}[f_1] = \mathbb{E}[\bar{f}_1]) \end{aligned}$$

Hence $f_1(\mathbf{y}_{A_1})$ and $f_2(\mathbf{y}_{A_2})$ are negatively correlated, so $\mathbf{y}_1, \dots, \mathbf{y}_n$ satisfy the definition of negative association. \square

10.2 Choosing W

Recall that we will set the output of \mathcal{W} to $(W(\mathbf{y}_1), W(\mathbf{y}_2), \dots)$ for $\mathbf{y} \sim \mathcal{D}$, where W is some increasing function. Our goal in this section is to make W as large as possible while still ensuring that the output of \mathcal{W} isn't too large, meaning Equation (3) holds.

First, consider the case when every A_j does not contain duplicates (meaning $r_j(i) \leq 1$ for all i). In this case, we have that the desired win probability of each vertex $i \in A_j$ is

$$\mathbf{w}_j(i) = \mathbf{x}_{k_j(i)} = W(\mathbf{y}_{k_j(i)})$$

One natural way to pick a winner this round is to pick the vertex with the largest $\mathbf{y}_{k_j(i)}$, breaking ties at random. Indeed, we choose W in such a way that this occurs. Let $\mathbf{i}_{\max}(y_1, \dots, y_m)$ be the function that return the index i maximizing y_i , choosing randomly from maximal y_i in the case of a tie. We define

$$W(y) := \Pr_{\mathbf{y}_2, \dots, \mathbf{y}_m \stackrel{\text{iid}}{\sim} \mathcal{D}^{(1)}} [\mathbf{i}_{\max}(y, \mathbf{y}_2, \dots, \mathbf{y}_m) = 1] \quad (35)$$

This $W(y)$ is a general choice that can work for any seed distribution \mathcal{D} and, as long as \mathcal{D} has negative correlations, this choice of W results in \mathcal{W} having negative correlations as well. Furthermore, the reason that we need Equation (3) is so that it's possible to design a tournament subroutine, and this choice of $W(y)$ allows for an easy tournament subroutine when each vertex only appears once in A_j : Just pick the vertex with largest desired win probability. Indeed, in Lemma 10.7, we show that Equation (3) holds in the case of $r = 1$, though it still needs to be verified separately for $r \geq 2$. First, we compute the function W for our \mathcal{D} .

Lemma 10.6. *For W defined in Equation (35) and \mathcal{D} in Definition 13,*

$$W(y) = \frac{G(y+1)^m - G(y)^m}{m \cdot (G(y+1) - G(y))} \quad \text{where} \quad G(v) := \begin{cases} 0 & \text{if } v = -1 \\ (1-p) + p \cdot (1 - (1-p)^v) & \text{if } v \geq 0 \end{cases} \quad (36)$$

Proof. Let $\mathbf{y}_1, \dots, \mathbf{y}_m$ be m independent samples from $\mathcal{D}^{(1)}$, and let $\mathbf{i}^* = \mathbf{i}_{\max}(\mathbf{y}_1, \mathbf{y}_2, \dots, \mathbf{y}_m)$. We wish to compute the probability $\mathbf{i}^* = 1$ as a function of \mathbf{y}_1 . First, we observe that G as defined in Equation (36) satisfies, for each $i = 1, \dots, m$,

$$G(v) = \Pr[\mathbf{y}_i \leq v - 1] \quad (37)$$

We consider a different process for generating $\mathbf{y}_1, \dots, \mathbf{y}_m$ that simplifies the analysis. In general, we can generate any univariate random variable \mathbf{y} by first generating $\mathbf{a} \sim \text{Uniform}(0, 1)$ and then setting \mathbf{y} to the smallest v such that $\Pr[\mathbf{y} \leq v] \geq \mathbf{a}$. Applying this to our setting, we first sample $\mathbf{a}_1, \dots, \mathbf{a}_m \sim \text{Uniform}(0, 1)$ and then for each $i \in [m]$ set \mathbf{y}_i to the smallest v such that $G(v+1) \geq \mathbf{a}_i$. Then, we set \mathbf{i}^* to the unique i maximizing \mathbf{a}_i (which exists with probability 1 since the \mathbf{a}_i are continuous).

This generating process for $(\mathbf{y}_1, \dots, \mathbf{y}_m, \mathbf{i}^*)$ gives the same joint distribution as the original, so we are free to analyze it instead. Then, the distribution of \mathbf{a}_i conditioned on \mathbf{y}_i is given by

$$(\mathbf{a}_i \mid \mathbf{y}_i = y) \sim \text{Uniform}(G(y), G(y+1)).$$

We next compute the desired result.

$$\begin{aligned}
\Pr[\mathbf{i}^* = 1 \mid \mathbf{y}_1 = y] &= \Pr[\mathbf{a}_i < \mathbf{a}_1 \text{ for all } i = 2, \dots, m \mid \mathbf{y}_1 = y] \\
&= \mathbb{E}_{\mathbf{a}_1} [(\mathbf{a}_1)^{m-1}] && (\mathbf{a}_i \sim \text{Uniform}(0, 1) \text{ for each } i = 2, \dots, m) \\
&= \frac{1}{G(y+1) - G(y)} \cdot \int_{G(y)}^{G(y+1)} t^{m-1} dt && (\mathbf{a}_1 \sim \text{Uniform}(G(y), G(y+1))) \\
&= \frac{G(y+1)^m - G(y)^m}{m \cdot (G(y+1) - G(y))},
\end{aligned}$$

which is exactly the definition of $W(y)$. □

Next, we prove that using the W from [Equation \(35\)](#) makes our win distribution satisfy [Equation \(3\)](#) for $r = 1$. We prove the following Lemma for our particular choice of \mathcal{D} but note that a similar result holds for any \mathcal{D} (as long as W is defined as [Equation \(35\)](#) as a function of that \mathcal{D}).

Lemma 10.7 (Sufficiently small for $r = 1$). *Let W be as defined in [Equation \(35\)](#). For any $t \in (0, 1)$,*

$$\mathbb{E}_{\mathbf{y} \sim \mathcal{D}^{(1)}} [(W(\mathbf{y}) - t)_+] \leq 1 - t - \frac{m-1}{m} \cdot \left(1 - t^{\frac{m}{m-1}}\right)$$

In particular, [Equation \(3\)](#) holds for $r = 1$.

Proof. Straightforward computation verifies that if $\mathbf{a} \sim \text{Uniform}(0, 1)$ and $\mathbf{z} = \mathbf{a}^{m-1}$, then for all $t \in (0, 1)$

$$\mathbb{E}_{\mathbf{z}} [(z - t)_+] = 1 - t - \frac{m-1}{m} \cdot \left(1 - t^{\frac{m}{m-1}}\right).$$

Let \mathbf{w} have the distribution of $W(\mathbf{y})$ when $\mathbf{y} \sim \mathcal{D}^{(1)}$. By [Fact 9.14](#), the desired result is equivalent to showing that $\mathbf{w} \preceq_{\text{icx}} \mathbf{z}$. We'll instead prove $\mathbf{w} \preceq_{\text{cx}} \mathbf{z}$ (a stronger statement) and do so via [Fact 8.2](#).

To use [Fact 8.2](#), we need to couple \mathbf{w} and \mathbf{z} . Consider the following coupled generating process.

1. Draw $\mathbf{a} \sim \text{Uniform}(0, 1)$.
2. Set $\hat{\mathbf{z}} = \mathbf{a}^{m-1}$.
3. Set $\hat{\mathbf{y}}$ to smallest $v \in \mathbb{Z}_{\geq -1}$ such that $G(v+1) \geq y$ where G is defined as in [Equation \(36\)](#).
4. Set $\hat{\mathbf{w}} = W(\hat{\mathbf{y}})$.

Clearly $\hat{\mathbf{z}} \stackrel{d}{=} \mathbf{z}$ as they have the same generative process. Furthermore, in the proof of [Lemma 10.6](#), we argued that $\hat{\mathbf{y}} \stackrel{d}{=} \mathbf{y}$ which also implies $\hat{\mathbf{w}} \stackrel{d}{=} \mathbf{w}$. Hence, in order to apply [Fact 8.2](#) and complete the proof of this Lemma, it is sufficient to show that $\mathbb{E}[\hat{\mathbf{x}} \mid \hat{\mathbf{w}}] = \hat{\mathbf{w}}$. Equivalently, we wish to show that

$$\mathbb{E}[\hat{\mathbf{x}} \mid \hat{\mathbf{y}}] = W(\hat{\mathbf{y}}).$$

Based on our generative process, the distribution $(\hat{\mathbf{a}} \mid \hat{\mathbf{y}})$ is given by $\text{Uniform}(G(\hat{\mathbf{y}}), G(\hat{\mathbf{y}} + 1))$. From the proof of [Lemma 10.6](#), we have that the expectation of $\hat{\mathbf{x}} = \mathbf{a}^{m-1}$ conditioned on $\mathbf{a} \sim \text{Uniform}(G(\hat{\mathbf{y}}), G(\hat{\mathbf{y}} + 1))$ is exactly given by $W(\hat{\mathbf{y}})$, as desired. □

For computational reasons, we don't directly use [Equation \(35\)](#). Instead, we use a ‘‘capped’’ version of W . For some $y_{\max} \geq 0$

$$W^{\text{capped}}(y) := W(\min(y, y_{\max})) - 10^{-5}.$$

For concreteness, $y_{\max} = 30$ is sufficient to compute our competitive ratio. Using $W^{\text{capped}}(\mathbf{y})$ instead of W in [Lemma 10.7](#) still works as it can only lead to a smaller $\mathbb{E}_{\mathbf{y} \sim \mathcal{D}(1)} \left[(W^{\text{capped}}(\mathbf{y}) - t)_+ \right]$. We have a y_{\max} to make the computation tractable. The 10^{-5} term is included because, without it, rounding errors make it difficult to determine if [Equation \(3\)](#) holds. Note that we also use W^{capped} in place of W when computing the competitive ratio, so with more accurate computation (and less aggressive capping), we could slightly increase the competitive ratio. We do not believe this would change any of the first 4 digits of accuracy which we report.

To summarize, our distribution \mathcal{W} is defined as follows: First draw $(\mathbf{y}_1, \mathbf{y}_2, \dots) \sim \mathcal{D}$, and then output $(\mathbf{x}_1, \mathbf{x}_2, \dots)$ where $\mathbf{x}_j = W^{\text{capped}}(\mathbf{y}_j)$. Since this W^{capped} is increasing, the output of \mathcal{W} is negatively associated. Furthermore, it is shift-invariant just as \mathcal{D} is.

We have verified that \mathcal{W} satisfies the first two criteria of [Theorem 4](#) and the third for $r = 1$. Lastly, we verify the third for $r = 2, \dots, m-1$ via a mix of straightforward algebra and computation. The first r elements of \mathcal{W} take on only finitely many values (though exponential in r). We fully compute a representation of this distribution for each $r = 2, \dots, m-1$. Given this representation, we are able to exactly compute the distribution for $\mathbf{w}^{(r)}$. Verifying [Equation \(3\)](#) amounts to verifying that $g(t) \geq 0$ for all $t \in [0, 1]$ where

$$g(t) := \left(1 - t - \frac{m-r}{m} \cdot \left(1 - t^{\frac{m}{m-r}} \right) \right) - \mathbb{E}_{\mathbf{x} \sim \mathcal{W}} \left[(\mathbf{w}^{(r)} - t)_+ \right].$$

Let S be the (finitely sized) support of $\mathbf{w}^{(r)}$. Then, the derivative of g is easy to compute.

$$g'(t) = \begin{cases} \Pr_{\mathbf{x} \sim \mathcal{W}} [\mathbf{w}^{(r)} \geq t] - 1 + t^{\frac{r}{m-r}} & \text{if } t \notin S \\ \text{undefined} & \text{otherwise} \end{cases}$$

We compute all the extreme points of g on the interval $[0, 1]$. If $g(t) < 0$ for some t on that interval, it will also be negative at some extreme point. The extreme points can only occur at the border ($t = 0$ or $t = 1$), where the derivative is undefined (on $t \in S$), or at a point where $g'(t) = 0$. Let V be the set of distinct values for $\Pr_{\mathbf{x} \sim \mathcal{W}} [\mathbf{w}^{(r)} \geq t]$, of which there are only $|S| + 1$ many. Then, if $g'(t) = 0$, it implies that, for some $v \in V$

$$t = (1 - v)^{\frac{m-r}{r}}.$$

Therefore, the total number of points we need to test for $g(t) < 0$ is finite (it has size $2 \cdot |S| + 3$). We verify that $g(t) \geq 0$ for each of those points in a publicly available Colab notebook³ implying that [Equation \(3\)](#) holds.

As demonstrated in the Colab notebook, for our choice of \mathcal{W} , [Equation \(3\)](#) doesn't hold for every choice of the hyperparameter p . We found that for $m = 6$, it holds for all $p \in \{0.05, 0.06, \dots, 0.95\}$, but not for $p = \{0.01, \dots, 0.04, 0.96, \dots, 0.99\}$. In particular, it holds for $p = 0.48$, which is the setting that maximized the resulting competitive ratio.

³To see our code, go to <https://colab.research.google.com/drive/1yQErphKVkwPPXsUWGT2b-nIPPaLBtNh?usp=sharing>

As a result of the above analysis and [Theorem 4](#), we have constructed an (F, m) -discrete OCS for the F defined in [Equation \(4\)](#), restated here for convenience.

$$F(n) := \mathbb{E}_{\mathbf{x} \sim \mathcal{W}} \left[\prod_{\ell=1}^n (1 - \mathbf{x}_\ell) \right]$$

Once again, we use computation (also in the Jupyter notebook) to determine $F(n)$. Up to $n_{\max} = 10$, we store a full representation of the first n_{\max} elements from \mathcal{W} and then use it to compute $F(n)$ for $n \leq n_{\max}$. For $n > n_{\max}$, we can upper bound $F(n) \leq F(n_{\max}) \cdot \left(\frac{F(n_{\max})}{F(n_{\max}-1)}\right)^{n-n_{\max}}$ as a result of the following Lemma.

Lemma 10.8. *For the function F defined in [Equation \(4\)](#), the quantity $\frac{F(n+1)}{F(n)}$ is decreasing in n .*

Our proof of [Lemma 10.8](#) holds for any increasing function W (so will also work for W^{capped}), but is specialized to our choice for \mathcal{D} from [Definition 13](#).

Proof. Draw $\mathbf{y}, \mathbf{z} \sim \mathcal{D}(y, z)$ and then set $\mathbf{a}_k = \text{Ber}(W(\mathbf{y}_k))$. We can think of \mathbf{a}_k as indicating whether the vertex i wins the k^{th} time it appears. Then the following equation for F is equivalent to [Equation \(4\)](#).

$$F(n) = \Pr[\mathbf{a}_1 = \dots = \mathbf{a}_n = 0]$$

Therefore,

$$\frac{F(n+1)}{F(n)} = \Pr[\mathbf{a}_{n+1} = 0 \mid \mathbf{a}_1 = \dots = \mathbf{a}_n = 0].$$

We want the above quantity to be less than or equal to

$$\frac{F(n)}{F(n-1)} = \Pr[\mathbf{a}_{n+1} = 0 \mid \mathbf{a}_2 = \dots = \mathbf{a}_n = 0].$$

using E to indicate the event $[\mathbf{a}_j = 0 \text{ for each } j \in \{2, \dots, n\}]$, proving that $\frac{F(n+1)}{F(n)} \leq \frac{F(n)}{F(n-1)}$ is equivalent to

$$\Pr[\mathbf{a}_{n+1} = 0 \mid E, \mathbf{a}_1 = 0] \leq \Pr[\mathbf{a}_{n+1} = 0 \mid E]. \tag{38}$$

For each $k = 1, \dots, n$, we define the stochastic function $\mathbf{h}_k : \mathbb{R}_{\geq 0} \rightarrow \mathbb{R}_{\geq 0}$,

$$\mathbf{h}_k(z) = (\mathbf{z}_{k+1} \mid \mathbf{z}_1 = z, \mathbf{a}_1 = \dots = \mathbf{a}_k = 0).$$

We claim that \mathbf{h}_k is a stochastically increasing function for all $k = 1, \dots, n$. For $k = 1$ this amounts to proving following map is stochastically increasing

$$z \mapsto (\mathbf{z}_2 \mid \mathbf{z}_1 = z, a_1 = 0).$$

Recall that by [Definition 13](#), \mathbf{z}_2 and \mathbf{z}_1 are related as follows.

1. With probability equal to p (the hyperparameter in [Definition 13](#)), $\mathbf{z}_2 = 0$. If this happens, then $\mathbf{y}_1 = \mathbf{z}_1$.

2. With probability $1 - p$, $z_2 = z_1 + 1$ and then $y_1 = -1$

If the probability that the first branch happens conditioned on $(z_1 = z, a_1 = 0)$ is decreasing in z then h_1 is stochastically increasing. We compute that probability using Bayes theorem.

$$\begin{aligned} \Pr[y_1 = z \mid a_1 = 0, z_1 = z] &= \frac{\Pr[a_1 = 0 \mid y_1 = z, z_1 = z] \cdot \Pr[y_1 = z \mid z_1 = z]}{\Pr[a_1 = 0 \mid z_1 = z]} \\ &= \frac{(1 - W(z))}{p \cdot (1 - W(z)) + (1 - p) \cdot (1 - W(-1))} \cdot \Pr[y_1 = z \mid z_1 = z] \end{aligned}$$

The quantity $\Pr[y_1 = z \mid z_1 = z]$ is a constant (it is just equal to p). Since W is increasing, the above is decreasing in z . Therefore, h_1 is stochastically decreasing.

For $k \geq 2$, by the inductive hypothesis, we may that assume h_{k-1} is stochastically increasing. Then,

$$\begin{aligned} h_k(z) &= (z_{k+1} \mid z_1 = z, \mathbf{a}_1 = \dots = \mathbf{a}_k = 0) \\ &= (z_{k+1} \mid z_2 = z', \mathbf{a}_2 = \dots = \mathbf{a}_k = 0) \quad \text{where } z' \sim (z_2 \mid z_1 = z, a_1 = 0) \quad (\text{Fact 10.1}) \\ &= (h_{k-1} \circ h_1)(z) \quad (\mathcal{D} \text{ is shift invariant}) \end{aligned}$$

and since both h_{k-1} and h_1 are stochastically increasing, so is h_k . We are now ready to prove [Equation \(38\)](#). If we condition on $\mathbf{a}_1 = 0$, then the distribution of z_2 is that of a $\text{Geo}(p) + 1$. If we do not condition on \mathbf{a}_1 , that distribution is just $\text{Geo}(p)$. Since h_n is stochastically increasing and $\text{Geo}(p) + 1$ stochastically dominates $\text{Geo}(p)$, the distribution of z_{n+1} conditioned on $\mathbf{a} = 0$ and E stochastically dominates the distribution of z_{n+1} just conditioning on E . The quantity y_{n+1} is a stochastically increasing function of z_{n+1} and the quantity \mathbf{a}_{n+1} is a stochastically increasing function of y_{n+1} . Therefore, the probability $\mathbf{a}_{n+1} = 0$ conditioned on E and $\mathbf{a}_1 = 0$ is lower than just conditioning on E . [Equation \(38\)](#) holds and therefore the desired result does as well. \square

As a consequence of the above, we have proven the existence of an $(F, 6)$ -OCS for the F in [Figure 5](#).

11 Proof of [Lemma 2.2](#)

Lemma 2.2. *No γ -OCS exists for $\gamma > \frac{1}{3}$.*

Proof. We prove that there is no 2-discrete OCS satisfying both of the following.

1. The probability a vertex is picked in a single time step it appears is $\frac{1}{2}$
2. Let j_1, j_2 be two time steps vertex i appears in for which i never appears in a time step between j_1 and j_2 . Then, the probability i is selected in j_1 or j_2 is strictly greater than $\frac{5}{6}$.

This implies [Lemma 2.2](#). Suppose, for the sake of contradiction, such an OCS exists. Feed in the following three pairs into the OCS. $((a, b), (b, c), (a, b))$. In the first round, the OCS must pick a and b each with probability $\frac{1}{2}$.

Then, in order to guarantee that b is picked in one of the first two time steps with probability more than $\frac{5}{6}$, if a is picked in the first round, b must be picked in the second round with probability $p_2 > \frac{2}{3}$.

We also need the probability that c is picked in the second round to be $\frac{1}{2}$. To guarantee this, in the case where b is picked in the first round, we need c to be picked with probability p_2 . Hence, there is a $\frac{1}{2} \cdot p_2$ probability that b is picked in the first round and c in the second. Let p_3 be the probability that a is picked in this case. If $p_3 \geq \frac{1}{2}$, then there is a

$$\frac{1}{2} \cdot p_2 \cdot p_3 > \frac{1}{6}$$

chance that b is not selected in the second or third round. On the other hand, if $p_3 \leq \frac{1}{2}$, then there is a

$$\frac{1}{2} \cdot p_2 \cdot (1 - p_3) > \frac{1}{6}$$

chance that a is not selected in the first or third round. In both cases, we have a contradiction. \square

References

- [AGKM11] Gagan Aggarwal, Gagan Goel, Chinmay Karande, and Aranyak Mehta. Online vertex-weighted bipartite matching and single-bid budgeted allocations. In *Proceedings of the twenty-second annual ACM-SIAM symposium on Discrete Algorithms*, pages 1253–1264. SIAM, 2011. [1](#), [5.1](#)
- [BJN07] Niv Buchbinder, Kamal Jain, and Joseph Seffi Naor. Online primal-dual algorithms for maximizing ad-auctions revenue. In *European Symposium on Algorithms*, pages 253–264. Springer, 2007. [1.1](#)
- [DH09] Nikhil R Devanur and Thomas P Hayes. The adwords problem: online keyword matching with budgeted bidders under random permutations. In *Proceedings of the 10th ACM conference on Electronic commerce*, pages 71–78, 2009. [1.1](#)
- [DHK⁺16] Nikhil R Devanur, Zhiyi Huang, Nitish Korula, Vahab S Mirrokni, and Qiqi Yan. Whole-page optimization and submodular welfare maximization with online bidders. *ACM Transactions on Economics and Computation (TEAC)*, 4(3):1–20, 2016. [7.1](#)
- [DJSW19] Nikhil R Devanur, Kamal Jain, Balasubramanian Sivan, and Christopher A Wilkens. Near optimal online algorithms and fast approximation algorithms for resource allocation problems. *Journal of the ACM (JACM)*, 66(1):1–41, 2019. [1.1](#)
- [FHK⁺10] Jon Feldman, Monika Henzinger, Nitish Korula, Vahab S Mirrokni, and Cliff Stein. Online stochastic packing applied to display ad allocation. In *European Symposium on Algorithms*, pages 182–194. Springer, 2010. [1.1](#)
- [FHTZ20] M. Fahrback, Z. Huang, R. Tao, and M. Zadimoghaddam. Edge-weighted online bipartite matching. In *2020 IEEE 61st Annual Symposium on Foundations of Computer Science (FOCS)*, pages 412–423, Los Alamitos, CA, USA, nov 2020. IEEE Computer Society. [1](#), [2.1](#), [1](#), [5](#), [2.2](#), [2.3](#), [4](#), [5.1](#), [7.1](#), [7.1](#)
- [FKM⁺09] Jon Feldman, Nitish Korula, Vahab Mirrokni, Shanmugavelayutham Muthukrishnan, and Martin Pál. Online ad assignment with free disposal. In *International workshop on internet and network economics*, pages 374–385. Springer, 2009. [1.1](#)

- [FMMM09] Jon Feldman, Aranyak Mehta, Vahab Mirrokni, and Shan Muthukrishnan. Online stochastic matching: Beating $1-1/e$. In *2009 50th Annual IEEE Symposium on Foundations of Computer Science*, pages 117–126. IEEE, 2009. [1.1](#)
- [GHH⁺21] Ruiquan Gao, Zhongtian He, Zhiyi Huang, Zipei Nie, Bijun Yuan, and Yan Zhong. Improved online correlated selection, 2021. [5.1](#)
- [GKM⁺19] Buddhima Gamlath, Michael Kapralov, Andreas Maggiori, Ola Svensson, and David Wajc. Online matching with general arrivals. In *2019 IEEE 60th Annual Symposium on Foundations of Computer Science (FOCS)*, pages 26–37. IEEE, 2019. [1.1](#)
- [GKS19] Buddhima Gamlath, Sagar Kale, and Ola Svensson. Beating greedy for stochastic bipartite matching. In *Proceedings of the Thirtieth Annual ACM-SIAM Symposium on Discrete Algorithms*, pages 2841–2854. SIAM, 2019. [1.1](#)
- [GM08] Gagan Goel and Aranyak Mehta. Online budgeted matching in random input models with applications to adwords. In *SODA*, volume 8, pages 982–991, 2008. [1.1](#)
- [HKT⁺18] Zhiyi Huang, Ning Kang, Zhihao Gavin Tang, Xiaowei Wu, Yuhao Zhang, and Xue Zhu. How to match when all vertices arrive online. In *Proceedings of the 50th Annual ACM SIGACT Symposium on Theory of computing*, pages 17–29, 2018. [1.1](#)
- [HMZ11] Bernhard Haeupler, Vahab S Mirrokni, and Morteza Zadimoghaddam. Online stochastic weighted matching: Improved approximation algorithms. In *International workshop on internet and network economics*, pages 170–181. Springer, 2011. [1.1](#)
- [HPT⁺19] Zhiyi Huang, Binghui Peng, Zhihao Gavin Tang, Runzhou Tao, Xiaowei Wu, and Yuhao Zhang. Tight competitive ratios of classic matching algorithms in the fully online model. In *Proceedings of the Thirtieth Annual ACM-SIAM Symposium on Discrete Algorithms*, pages 2875–2886. SIAM, 2019. [1.1](#)
- [HTWZ19] Zhiyi Huang, Zhihao Gavin Tang, Xiaowei Wu, and Yuhao Zhang. Online vertex-weighted bipartite matching: Beating $1-1/e$ with random arrivals. *ACM Transactions on Algorithms (TALG)*, 15(3):1–15, 2019. [1.1](#)
- [JDP83] Kumar Joag-Dev and Frank Proschan. Negative Association of Random Variables with Applications. *The Annals of Statistics*, 11(1):286 – 295, 1983. [6](#), [3](#), [9.1](#), [6](#)
- [JL14] Patrick Jaillet and Xin Lu. Online stochastic matching: New algorithms with better bounds. *Mathematics of Operations Research*, 39(3):624–646, 2014. [1.1](#)
- [KMT11] Chinmay Karande, Aranyak Mehta, and Pushkar Tripathi. Online bipartite matching with unknown distributions. In *Proceedings of the forty-third annual ACM symposium on Theory of computing*, pages 587–596, 2011. [1.1](#)
- [KP00] Bala Kalyanasundaram and Kirk R Pruhs. An optimal deterministic algorithm for online b-matching. *Theoretical Computer Science*, 233(1-2):319–325, 2000. [1.1](#)
- [KVV90] Richard M Karp, Umesh V Vazirani, and Vijay V Vazirani. An optimal algorithm for on-line bipartite matching. In *Proceedings of the twenty-second annual ACM symposium on Theory of computing*, pages 352–358, 1990. [1](#)

- [Meh13] Aranyak Mehta. Online matching and ad allocation. *Foundations and Trends® in Theoretical Computer Science*, 8(4):265–368, 2013. [1.1](#)
- [MGS12] Vahideh H Manshadi, Shayan Oveis Gharan, and Amin Saberi. Online stochastic matching: Online actions based on offline statistics. *Mathematics of Operations Research*, 37(4):559–573, 2012. [1.1](#)
- [MGZ12] Vahab S Mirrokni, Shayan Oveis Gharan, and Morteza Zadimoghaddam. Simultaneous approximations for adversarial and stochastic online budgeted allocation. In *Proceedings of the twenty-third annual ACM-SIAM symposium on Discrete Algorithms*, pages 1690–1701. SIAM, 2012. [1.1](#)
- [MP12] Aranyak Mehta and Debmalya Panigrahi. Online matching with stochastic rewards. In *2012 IEEE 53rd Annual Symposium on Foundations of Computer Science*, pages 728–737. IEEE, 2012. [1.1](#)
- [MR01] Alfred Müller and Ludger Rüschendorf. On the optimal stopping values induced by general dependence structures. *Journal of applied probability*, pages 672–684, 2001. [3](#), [8](#)
- [MSVV07] Aranyak Mehta, Amin Saberi, Umesh Vazirani, and Vijay Vazirani. Adwords and generalized online matching. *Journal of the ACM (JACM)*, 54(5):22–es, 2007. [1.1](#)
- [MWZ14] Aranyak Mehta, Bo Waggoner, and Morteza Zadimoghaddam. Online stochastic matching with unequal probabilities. In *Proceedings of the twenty-sixth annual ACM-SIAM symposium on Discrete algorithms*, pages 1388–1404. SIAM, 2014. [1.1](#)
- [MY11] Mohammad Mahdian and Qiqi Yan. Online bipartite matching with random arrivals: an approach based on strongly factor-revealing lps. In *Proceedings of the forty-third annual ACM symposium on Theory of computing*, pages 597–606, 2011. [1.1](#)
- [SA21] Yongho Shin and Hyung-Chan An. Making three out of two: Three-way online correlated selection. *CoRR*, abs/2107.02605, 2021. [5.2](#)
- [SS07] Moshe Shaked and J George Shanthikumar. *Stochastic orders*. Springer Science & Business Media, 2007. [6](#), [8.2](#), [8.3](#), [9.14](#)
- [Str65] Volker Strassen. The existence of probability measures with given marginals. *Annals of Mathematical Statistics*, 36(2):423–439, 1965. [3](#), [9.2](#)

A Missing proofs from [Section 2](#)

A.1 Proof that independent randomness is not enough

We prove that algorithms which make independent selections at each time step cannot achieve a competitive ratio better than that of the greedy algorithm.

Lemma 2.1. *Even for unweighted online bipartite matching, no algorithm making decisions that are independent across time steps can achieve a competitive ratio better than 0.5.*

Proof. For each $n \in \mathbb{Z}_{\geq 1}$, we'll show that no algorithm using independent randomness can achieve a competitive ratio better than $\frac{1}{2} + \frac{1}{2n}$ on every graph with n offline vertices. Fix some $n \in \mathbb{Z}_{\geq 1}$ and let \mathcal{G} be the family of triangular graphs with vertices $L = R = [n]$ defined as follows: For each permutation π of $[n]$, there is one graph $G \in \mathcal{G}$. The first online vertex has an edge to every offline vertex. Then, for each online vertex where $j \geq 2$ has an edge to every offline vertex that $j - 1 \in R$ has an edge to *except for* π_{j-1} .

It's not too hard to see that each $G \in \mathcal{G}$ has a perfect matching of all n edges formed by matching j to π_j . We'll prove that for any algorithm using independent randomness, there is some $G \in \mathcal{G}$ on which that algorithm has an expected number of matches at most $\frac{n+1}{2}$.

Let \mathcal{A} be some algorithm that uses independent randomness. At time step j , as a deterministic function of the portion of the graph revealed so far, the algorithm decides on a distribution over L on which to match the j^{th} vertex to. Then, it independently picks a match for j independent of all previous matches. In particular, if it picks a vertex in L that has been picked by some $j' < j$, this step is wasted. We'll use $p^{(j)}$ to represent the distribution over L that \mathcal{A} chooses for $j \in R$, where $(p^{(j)})_i$ is the probability that j is matched to i .

We will adversarial choose a $G \in \mathcal{G}$ depending on the algorithm \mathcal{A} . Once the j^{th} vertex arrives, the edges of each $j' \in R$ for $j' \leq j$ are known. This corresponds to knowing π_1, \dots, π_{j-1} . Then, \mathcal{A} must choose $p^{(j)}$, and we are free to adversarially choose π_j as it only affects the portion of the graph not yet revealed. Recall that whichever vertex we select as π_j will be unable to be matched in the future, as it will not have edges to any $j' \in R$ where $j' > j$. We'll set π_j to the offline vertex not yet removed that is *least likely* to be already matched. Formally,

$$\pi_j = \arg \min_{i \in L, i \neq \pi_{j'} \text{ for any } j' < j} \left(1 - \prod_{j' \leq j} (1 - (p^{(j')})_i) \right). \quad (39)$$

Next, we show that for any \mathcal{A} using independent randomness, that on this adversarial choice of $G \in \mathcal{G}$, \mathcal{A} matches at most $\frac{n+1}{2}$ edges in expectation. For each $k \in [n]$, we define R_k to be the sum over the last k offline nodes to be removed (the nodes $\pi_{n-k+1}, \dots, \pi_n$), that each is *not* matched to one of the first $n - k$ online nodes. Formally,

$$R_k := \sum_{i \in \{\pi_{n-k+1}, \dots, \pi_n\}} [i \text{ is not matched to } j \in \{\pi_1, \dots, \pi_{n-k}\}]$$

We claim, and will prove by induction, that the expected number of matches made to the last k offline nodes (the nodes $\pi_{n-k+1}, \dots, \pi_n$) is at most $k - R_k \cdot \frac{k-1}{2k}$. As $R_n = n$, this immediately implies the expected number of matches in the entire graph is at most $\frac{n+1}{2}$, the desired result.

For our base case, when $k = 1$, we trivially have the expected number of matches of the last vertex is at most 1. For the inductive case, fix some $k \geq 2$. We consider the state after $n - k$ rounds have occurred. For each $\ell \in [k]$, let r_ℓ be the probability that $\pi_{n-\ell+1}$ is not matched in any previous round (to $j \in [n - k]$) and $q_\ell := 1 - (p^{(j)})_{\pi_{n-\ell+1}}$ be the probability it is not matched in the next round (to $j = n - k + 1$), given that it was not matched in previous rounds. As adversarially chosen in [Equation \(39\)](#), π_{n-k+1} will be set to whichever node maximizes $r_\ell \cdot q_\ell$. Let ℓ^* be the ℓ maximizing $r_\ell \cdot q_\ell$. Then, $R_{k-1} = \sum_{\ell \neq \ell^*} r_\ell \cdot q_\ell$. The expected number of matches to the last k

offline vertices is at most

$$f(r, q) := \underbrace{1 - r_{\ell^*} q_{\ell^*}}_{\Pr[\pi_{n-k+1} \text{ matched}]} + \underbrace{k - 1 - \frac{k-2}{2(k-1)} \cdot \sum_{\ell \neq \ell^*} r_\ell \cdot q_\ell}_{\mathbb{E}[\text{matches to } \pi_{n-k+2}, \dots, \pi_n]} \quad \text{where } \ell^* = \arg \max_{\ell \in [k]} r_\ell \cdot q_\ell.$$

We used the inductive hypothesis to bound $\mathbb{E}[\text{matches to } \pi_{n-k+2}, \dots, \pi_n]$. Furthermore, r_ℓ and q_ℓ obey some bounds

$$\begin{aligned} \sum_{\ell \in [k]} r_\ell &= R_k && \text{(Definition of } R_k) \\ \sum_{\ell \in [k]} q_\ell &= \sum_{\ell \in [k]} 1 - (p^{(j)})_{\pi_{n-\ell+1}} = k - 1 && (40) \\ r_\ell, q_\ell &\geq 0 && \text{(for all } \ell \in [k]) \end{aligned}$$

The constraints of [Equation \(40\)](#) cause r_1, \dots, r_k and q_1, \dots, q_k to lie in a closed and bounded space, and f is continuous. Therefore, f attains its maximum value in this space. We may assume that $R_k > 0$, as if $R_k = 0$, f is $k - 1$ everywhere satisfying the above constraints. Similarly, as we only care about cases when $k \geq 2$, the sum of q_ℓ is also strictly more than 0. We claim the maximum of f subject to constraints of [Equation \(40\)](#) occurs when

$$r_1 = \dots = r_k = \frac{R_k}{k} \quad \text{and} \quad q_1 = \dots = q_k = \frac{k-1}{k}. \quad (41)$$

Consider any other point, $(r_1, \dots, r_k, q_1, \dots, q_k)$, satisfying the constraints of [Equation \(40\)](#). Using the notation $\frac{\partial f}{\partial x_+}$ to refer to the *right* derivative, how much f changes as x is increased by an ε amount, and $\frac{\partial f}{\partial x_-}$ for the left derivatives,

1. If $r_\ell \cdot q_\ell < r_{\ell'} \cdot q_{\ell'}$ for all $\ell' \neq \ell$, then,

$$\frac{\partial f}{\partial r_\ell} = -\frac{k-2}{2(k-1)} \cdot q_i \quad \frac{\partial f}{\partial q_\ell} = -\frac{k-2}{2(k-1)} \cdot r_i.$$

We note that the term $\frac{k-2}{2(k-1)}$ is in $[0, \frac{1}{2}]$ for $k \geq 2$. In particular, it is never more than 1, and strictly more than 0 whenever $k \geq 3$. When $k = 2$, it is easy to show the maximum of f subject to the constraints of [Equation \(40\)](#) occurs at the point in [Equation \(41\)](#). For the remainder of this proof, we consider the $k \geq 3$ case where $0 < \frac{k-2}{2(k-1)} < \frac{1}{2}$.

2. If $r_\ell \cdot q_\ell > r_{\ell'} \cdot q_{\ell'}$ for all $\ell' \neq \ell$, then,

$$\frac{\partial f}{\partial r_\ell} = -q_i \quad \frac{\partial f}{\partial q_\ell} = -r_i.$$

3. If $r_\ell \cdot q_\ell \geq r_{\ell'} \cdot q_{\ell'}$ for all $\ell' \neq \ell$, and there is at least one $\ell' \neq \ell$ where $r_\ell \cdot q_\ell = r_{\ell'} \cdot q_{\ell'}$, then

$$\begin{aligned} \frac{\partial f}{\partial (r_\ell)_+} &= -q_i && \frac{\partial f}{\partial (q_\ell)_+} &= -r_i. \\ \frac{\partial f}{\partial (r_\ell)_-} &= -\frac{k-2}{2(k-1)} \cdot q_i && \frac{\partial f}{\partial (q_\ell)_-} &= -\frac{k-2}{2(k-1)} \cdot r_i. \end{aligned}$$

Consider some point $(r_1, \dots, r_k, q_1, \dots, q_k)$ satisfying the constraints of Equation (40) that is not the point in Equation (41). Then, there exists some a, b where $r_a \cdot q_a \geq r_\ell \cdot q_\ell$ for each $\ell \neq a$, and either $r_a \neq r_b$ or $q_a \neq q_b$. Then, either $r_a > r_b$ or $q_a > q_b$. Without loss of generality, we assume $r_a > r_b$. Then, for sufficiently small ε , if we decrease $q_a \leftarrow q_a - \varepsilon$ and increase $q_b \leftarrow q_b + \varepsilon$, the value of f will increase and the constraints remain satisfied. Hence, the maximum value of f is

$$\begin{aligned} \max_{r,q} f(r, q) &= 1 - \frac{R_k}{k} \frac{k-1}{k} + k - 1 - \frac{k-2}{2(k-1)} \cdot (k-1) \cdot \frac{R_k}{k} \cdot \frac{k-1}{k} \\ &= k - R_k \cdot \frac{k-1}{2k}. \end{aligned}$$

Therefore, by induction, the maximum expected matches made to the last k offline nodes is at most $k - R_k \cdot \frac{k-1}{2k}$. Plugging in $k = n$ gives the desired result. \square

A.2 Proof that it is impossible to negatively correlate all time steps

We prove that there does not exist a continuous OCS that negatively correlates every single time step. This is why we choose to have our OCS only negatively correlate time steps that are close.

Lemma A.1. *Let $f : \mathbb{R}_{\geq 0} \rightarrow [0, 1]$ be a function on which there exists some rational $x \in \mathbb{Q}$ where $f(x) < e^{-x}$. Then, there does not exist a continuous OCS with the following guarantee: For any element $i \in L$ and set of time steps S ,*

$$\Pr[i_j \neq i \text{ for every } j \in S] \leq f\left(\sum_{j \in S} (p^{(j)})_i\right).$$

Note that $f(x) = e^{-x}$ corresponds to an OCS which picks an winner independently at each time step, and negative correlation corresponds to $f(x) < e^{-x}$.

Proof. Let $x = \frac{a}{b}$ for $a, b \in \mathbb{Z}$. For $k, n \in \mathbb{Z}_{\geq 1}$, consider the input instance where $|L| = b \cdot k$ and each of n time steps have $p^{(j)}$ uniform over L , meaning $(p^{(j)})_i = \frac{1}{b \cdot k}$ for all $i \in L, j \in [n]$. We will show that in the limit of $\frac{n}{k} \rightarrow \infty$ and $k \rightarrow \infty$, the desired guarantee is impossible.

Assume, for the sake of guarantee, the desired result is possible. We will show that this implies that the expected number of total winners is more than n , a contradiction as each time step has only a single winner. Fix an arbitrary $i \in V$, and let z_j be the event that i is *not* the winner in the j^{th} round. Then, by Jensen's inequality,

$$\mathbb{E}\left[\sum_{j \in [n]} z_j\right] \leq \sqrt{ak} \mathbb{E}\left[\left(\sum_{j \in [n]} z_j\right)^{ak}\right].$$

Expanding $(\sum_{j \in [n]} z_j)^{ak}$, there are $\frac{n!}{(n-ak)!}$ terms which are the product of ak distinct $z_{j_1}, \dots, z_{j_{ak}}$, and $n^{ak} - \frac{n!}{(n-ak)!}$ which are the product of non-distinct events. Whenever z_{j_1}, \dots, z_{j_k} are non-distinct, we use the loose upper bound that $E[z_{j_1} \cdots z_{j_{ak}}] \leq 1$. Whenever they are distinct,

by the assumption on the quality of the OCS, it must be the case that $E[z_{j_1} \cdots z_{j_{ak}}] \leq f(x)$. Therefore,

$$\begin{aligned} \mathbb{E} \left[\sum_{j \in [n]} z_j \right] &\leq \sqrt[ak]{\frac{n!}{(n-ak)!} \cdot f(x) + n^{ak} - \frac{n!}{(n-ak)!}} \\ &\leq \sqrt[ak]{\frac{n!}{(n-ak)!}} \cdot \sqrt[ak]{f(x)} + \sqrt[ak]{n^{ak} - \frac{n!}{(n-ak)!}} \end{aligned}$$

In the limit of $\frac{n}{k} \rightarrow \infty$,

$$\begin{aligned} \mathbb{E} \left[\sum_{j \in [n]} z_j \right] &\leq n \cdot (1 - o(1)) \cdot \sqrt[ak]{f(x)} \\ &< n \cdot (1 - o(1)) \cdot \sqrt[ak]{e^{-\frac{ak}{bk}}} \\ &= n \cdot (1 - o(1)) \cdot e^{-\frac{1}{bk}} \end{aligned}$$

Note that the expected number of times that i is the winner is $n - \mathbb{E} \left[\sum_{j \in [n]} z_j \right]$. In the limit of $k \rightarrow \infty$, this value is strictly more than $n \cdot \frac{1}{bk}$. By symmetry, this is true of all bk different vertices in L , implying, the total number of winner is more than n , a contradiction. \square

B Missing proofs from Section 7

Lemma 7.2. *Let $f : \mathbb{R}_{\geq 0} \rightarrow [0, 1]$ be thrice differentiable, convex, log concave, have a log concave derivative, and satisfy $f(0) = 1$. Define $a : \mathbb{R}_{\geq 0} \rightarrow [0, \Gamma]$ as*

$$a(x) := f(x) - \int_0^\infty e^{-t} f(t+x) dt. \quad (11)$$

For r satisfying Equation (9) and Γ defined in Equation (8), a satisfies the below properties.

$$a'(x) = f'(x) + a(x) \quad \forall x \geq 0 \quad (12)$$

$$a'(x) \leq 0 \quad \forall x \geq 0 \quad (13)$$

$$a(0) = \Gamma \quad (14)$$

$$f(x) \cdot a(0) = f(x) \cdot \Gamma \leq a(x) \quad \forall x \geq 0 \quad (15)$$

$$r \cdot (a(x) - f(x) \cdot \Gamma) \leq \Gamma - a(x) \quad \forall x \geq 0 \quad (16)$$

Proof of Equation (12). Using integration by parts,

$$\begin{aligned} a'(x) &= f'(x) - \int_0^\infty e^{-t} f'(t+x) dt \\ a'(x) &= f'(x) - \left(\left[e^{-t} f(t+x) \right]_{t=0}^{t=\infty} + \int_0^\infty e^{-t} f(t+x) dt \right) \quad (\text{integration by parts}) \\ a'(x) &= f'(x) - \left(-f(x) + \int_0^\infty e^{-t} f(t+x) dt \right) \\ &= f'(x) + a(x). \end{aligned}$$

□

Proof of Equation (13). Using the convexity of f ,

$$\begin{aligned} a'(x) &= f'(x) - \int_0^\infty e^{-t} f'(t+x) \\ &\leq f'(x) - \int_0^\infty e^{-t} f'(x) && (f \text{ is convex}) \\ &= 0 \end{aligned}$$

□

Proof of Equation (14). This is immediate from the fact that $f(0) = 1$ and the definitions of a in Equation (11) and of Γ in Equation (8). □

We'll use the following proposition.

Proposition B.1. *If $f : \mathbb{R}_{\geq 0} \rightarrow \mathbb{R}_{\geq 0}$ is log-concave, then*

$$\int_0^\infty \frac{e^{-t} f(t+x)}{f(x)} dt$$

is decreasing in x .

Proof. It is enough for $\frac{f(t+x)}{f(x)}$ to be decreasing in x for any fixed $t \geq 0$. The derivative of this expression is

$$\frac{d}{dx} \left(\frac{f(t+x)}{f(x)} \right) = \frac{f(x)f'(t+x) - f(t+x)f'(x)}{f(x)^2}.$$

Hence, the desired expression is decreasing when

$$\frac{f'(x)}{f(x)} \geq \frac{f'(t+x)}{f(t+x)}.$$

f is log-concave iff $\frac{f'(x)}{f(x)}$ is decreasing in x . Therefore, the desired result follows by the fact that $t \geq 0$ and the log-concavity of f . □

Proof of Equation (15). If $f(x) = 0$, the desired result holds. We wish to show that, for all $x \geq 0$ where $f(x) > 0$,

$$\frac{f(x) \cdot \Gamma}{a(x)} \leq 1.$$

As $f(0) = 1$ and $a(0) = \Gamma$, the above expression is 1 when $x = 0$. Therefore, it is enough to show that expression is decreasing. That's equivalent to showing that $\frac{f(x)}{a(x)}$ is decreasing, which, since we only need to worry about the case where $f(x) > 0$, is equivalent to showing that $\frac{a(x)}{f(x)}$ is increasing.

$$\begin{aligned} \frac{a(x)}{f(x)} &= \frac{f(x) - \int_0^\infty e^{-t} f(t+x) dt}{f(x)} \\ &= 1 - \int_0^\infty \frac{e^{-t} f(t+x)}{f(x)} dt \end{aligned}$$

By Proposition B.1, the above is increasing as a consequence of the log-concavity of f . □

In order to prove [Equation \(16\)](#), we'll need the following proposition.

Proposition B.2. *For any continuous $f, g : \mathbb{R}_{\geq 0} \rightarrow \mathbb{R}_{\geq 0}$ where $g > 0$ satisfying*

$$\frac{f(x)}{g(x)} \quad \text{is decreasing for all } x \geq 0.$$

Let $F(x) := \int_0^x f(t)dt$ and $G(x) := \int_0^x g(t)dt$. Then, the following is also decreasing for all $x \geq 0$,

$$\frac{F(x)}{G(x)}.$$

Proof. We compute the derivative of the desired expression:

$$\frac{d}{dx} \frac{F(x)}{G(x)} = \frac{G(x)f(x) - F(x)g(x)}{G(x)^2}$$

With a bit of rearranging, we have that the following are equivalent for all $x > 0$.

$$\begin{aligned} \frac{d}{dx} \frac{F(x)}{G(x)} < 0 & \iff z(x) := \frac{f(x)}{g(x)} - \frac{F(x)}{G(x)} < 0 \\ \frac{d}{dx} \frac{F(x)}{G(x)} = 0 & \iff z(x) := \frac{f(x)}{g(x)} - \frac{F(x)}{G(x)} = 0 \\ \frac{d}{dx} \frac{F(x)}{G(x)} > 0 & \iff z(x) := \frac{f(x)}{g(x)} - \frac{F(x)}{G(x)} > 0 \end{aligned}$$

Also, by L' Hôpital's rule, $\lim_{x \rightarrow 0} z(x) = 0$. By continuity of z , if $z(x) > 0$ for some $x > 0$, it must imply for some earlier $x^* \in (0, x)$ that $z(x^*) \geq 0$ and $z'(x^*) > 0$. However, as $\frac{f(x)}{g(x)}$ is decreasing, if $z'(x^*) > 0$, it must be the case that $\frac{F(x)}{G(x)}$ is decreasing at x^* . This contradicts that $z(x^*) \geq 0$. Therefore, $z(x) \leq 0$ for all $x > 0$ and the desired expression is decreasing. \square

Proof of [Equation \(16\)](#). Combining [Equations \(13\)](#) and [\(14\)](#), we have that $a(x) \leq \Gamma$ for all $x \geq 0$. Therefore, if $r = 0$, we are done. Otherwise, we wish to show that, for all $x \geq 0$,

$$\frac{1}{r} \geq h(x) := \frac{a(x) - f(x) \cdot \Gamma}{\Gamma - a(x)}$$

We will prove that $h(x)$ is decreasing. Assuming that is true, the desired result follows because

$$\begin{aligned} h(x) &\leq \lim_{x \rightarrow 0} h(x) = \frac{a'(0) - f'(0) \cdot \Gamma}{-a'(0)} && \text{(L' Hôpital's rule)} \\ &= \frac{f'(0) + a(0) - f'(0) \cdot \Gamma}{-(f'(0) + a(0))} && \text{(Equation (12))} \\ &= \frac{\Gamma + (1 - \Gamma) \cdot f'(0)}{-f'(0) - \Gamma} && \text{(Equation (14))} \\ &\leq \frac{1}{r} && \text{(Equation (9))} \end{aligned}$$

Let us define $A(x) := \frac{a(x)}{\Gamma}$. Then

$$\begin{aligned}
h(x) &= \frac{A(x) - f(x)}{1 - A(x)} \\
&= \frac{(1 - f(x)) - (1 - A(x))}{1 - A(x)} \\
&= \frac{1 - f(x)}{1 - A(x)} - 1 \\
&= \frac{\int_0^x f'(t) dt}{\int_0^x A'(t) dt} - 1 \qquad (f(0) = A(0) = 1)
\end{aligned}$$

By [Proposition B.2](#), in order to show $h(x)$ is decreasing, it is enough to show that $\frac{f'(x)}{A'(x)}$ is decreasing. Recall that $A(x) := \frac{a(x)}{\Gamma}$, so equivalently, we can show that $\frac{f'(x)}{a'(x)}$ is decreasing. Using the fact that $a'(x)$ and $f'(x)$ never change signs, we can instead show that $\frac{a'(x)}{f'(x)}$ is increasing.

$$\begin{aligned}
\frac{a'(x)}{f'(x)} &= \frac{f'(x) - \int_0^\infty e^{-t} f'(t+x) dt}{f'(x)} \qquad (\text{Equation (11)}) \\
&= 1 - \int_0^\infty \frac{e^{-t} f'(t+x)}{f'(x)} dt
\end{aligned}$$

By [Proposition B.1](#) and the log concavity of f' , the above is increasing, completing the proof of [Equation \(16\)](#). □