Online Correlated Selection

Zhiyi Huang
University of Hong Kong

Tutorial at COCOON 2024

Today’s Agenda

Part I

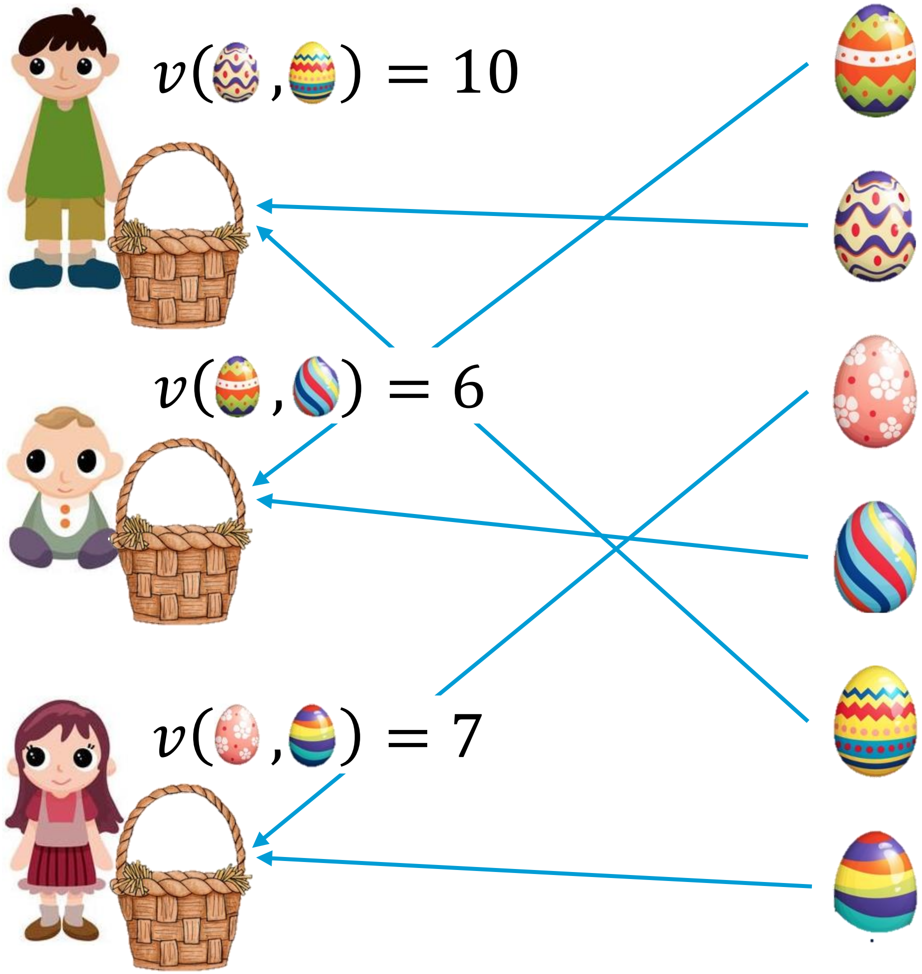
- Intro: What is Online Correlated Selection?
- Optimal two-way OCS for matching
- OCS for Display Ads

Part II

- Multi-way OCS for matching
- Optimal two-way Stochastic OCS for matching
- Outro: Summary, directions, and connections to other problems

Online Submodular Welfare Maximization

Kapralov, Post, and Vondrák (2013)

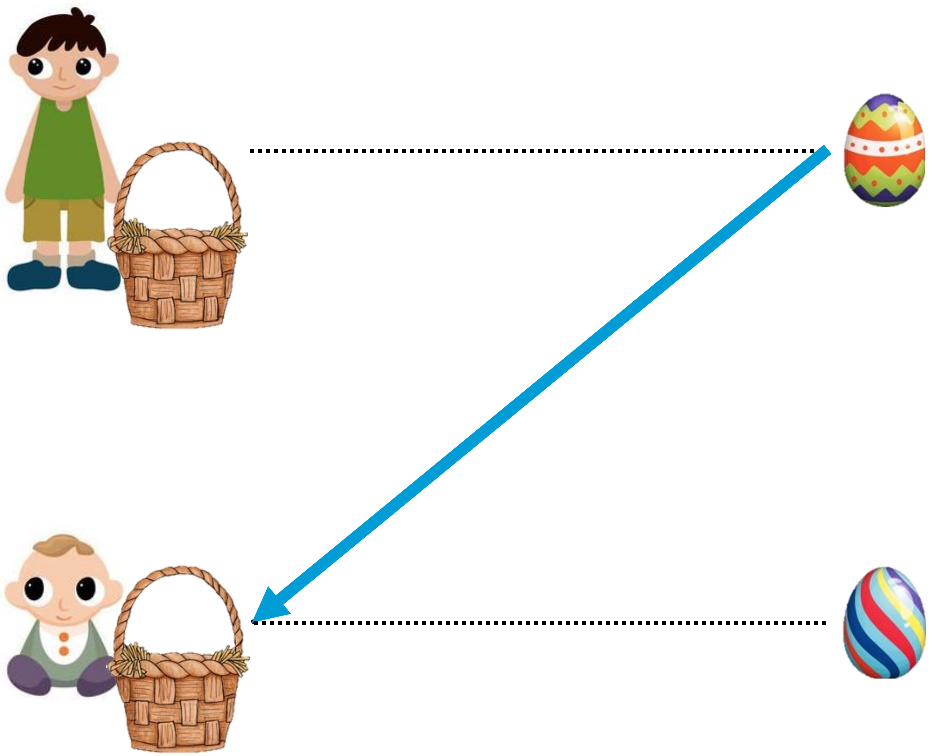


- n agents
  submodular values over subsets of items
- m items arriving online
  allocate each one as it comes
- Maximize welfare
  i.e., sum of agents' values
- Competitive analysis

$$\frac{\text{ALG}}{\text{OPT}}$$

Online Bipartite Matching

Karp, Vazirani, and Vazirani (1990)



$$v(S) = \begin{cases} 1 & \text{has a neighbor in } S \\ 0 & \text{otherwise} \end{cases}$$

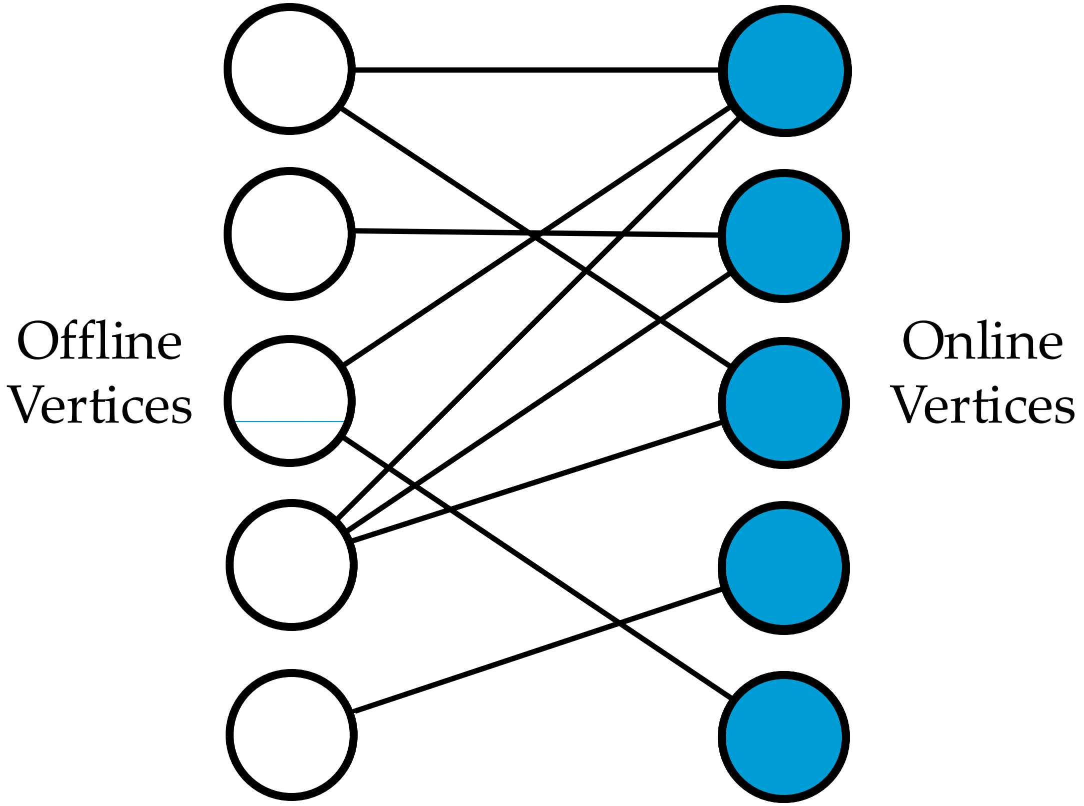
Don't put all your eggs in one basket!

Balance Algorithm for Fractional Matching

Kalyanasundaram and Pruhs (2000)

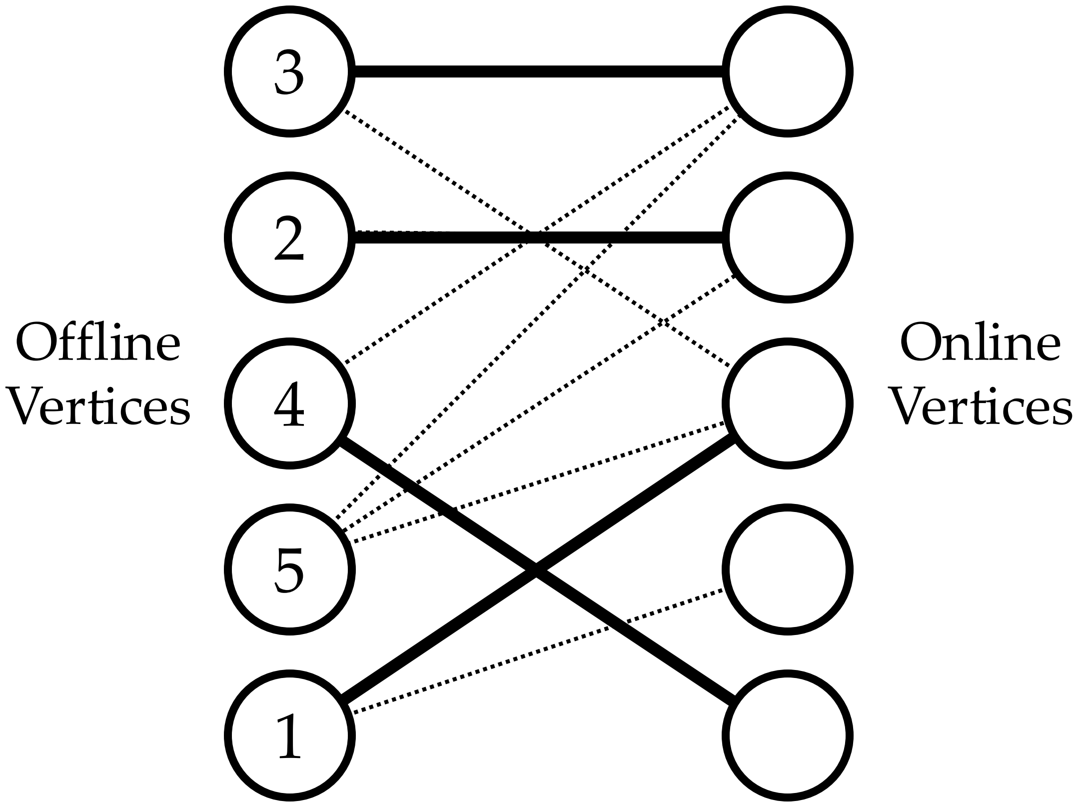
- Match each item (online vertex) to least matched neighbors
- "Water level" of an agent (offline vertex) is its matched portion

Balance is $1 - {}^{1}/_{e}$ competitive, which is optimal



Ranking Algorithm

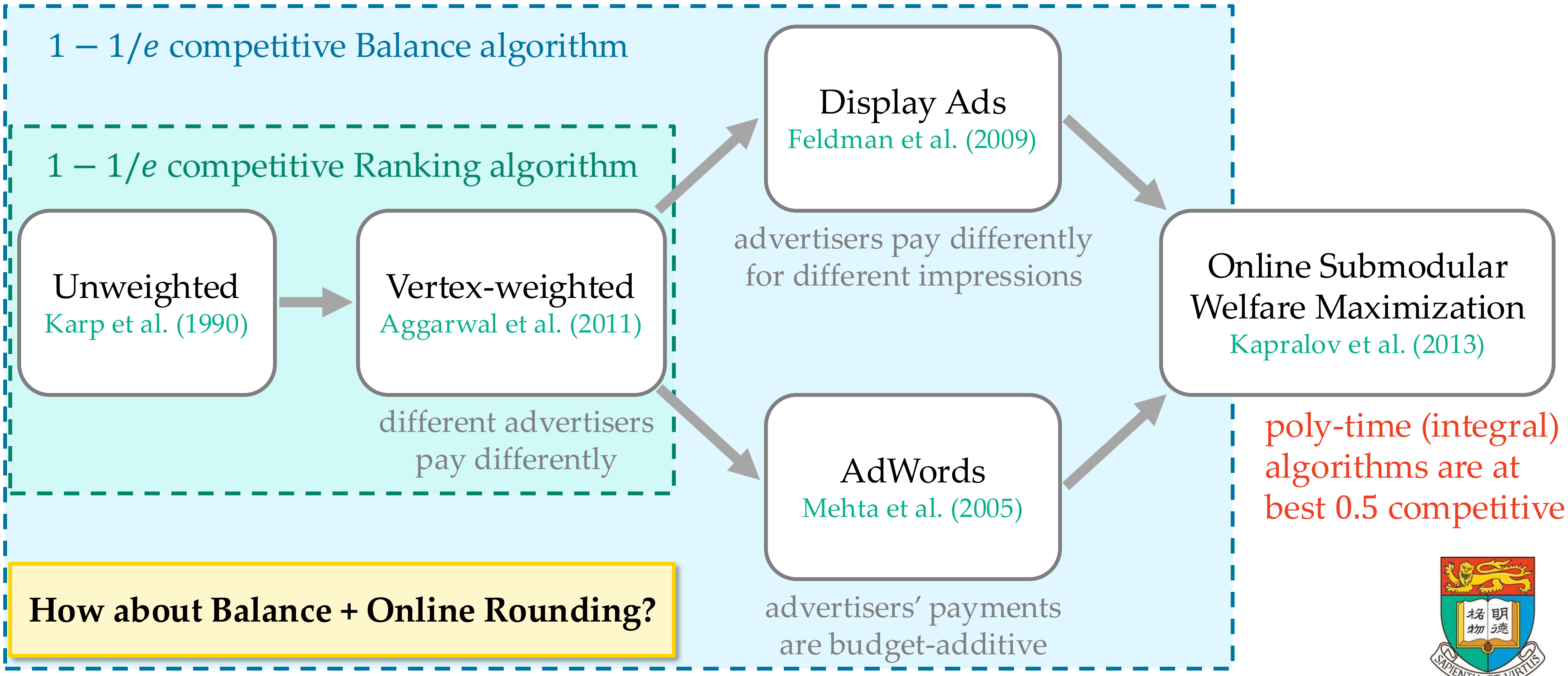
Karp, Vazirani, and Vazirani (1990)



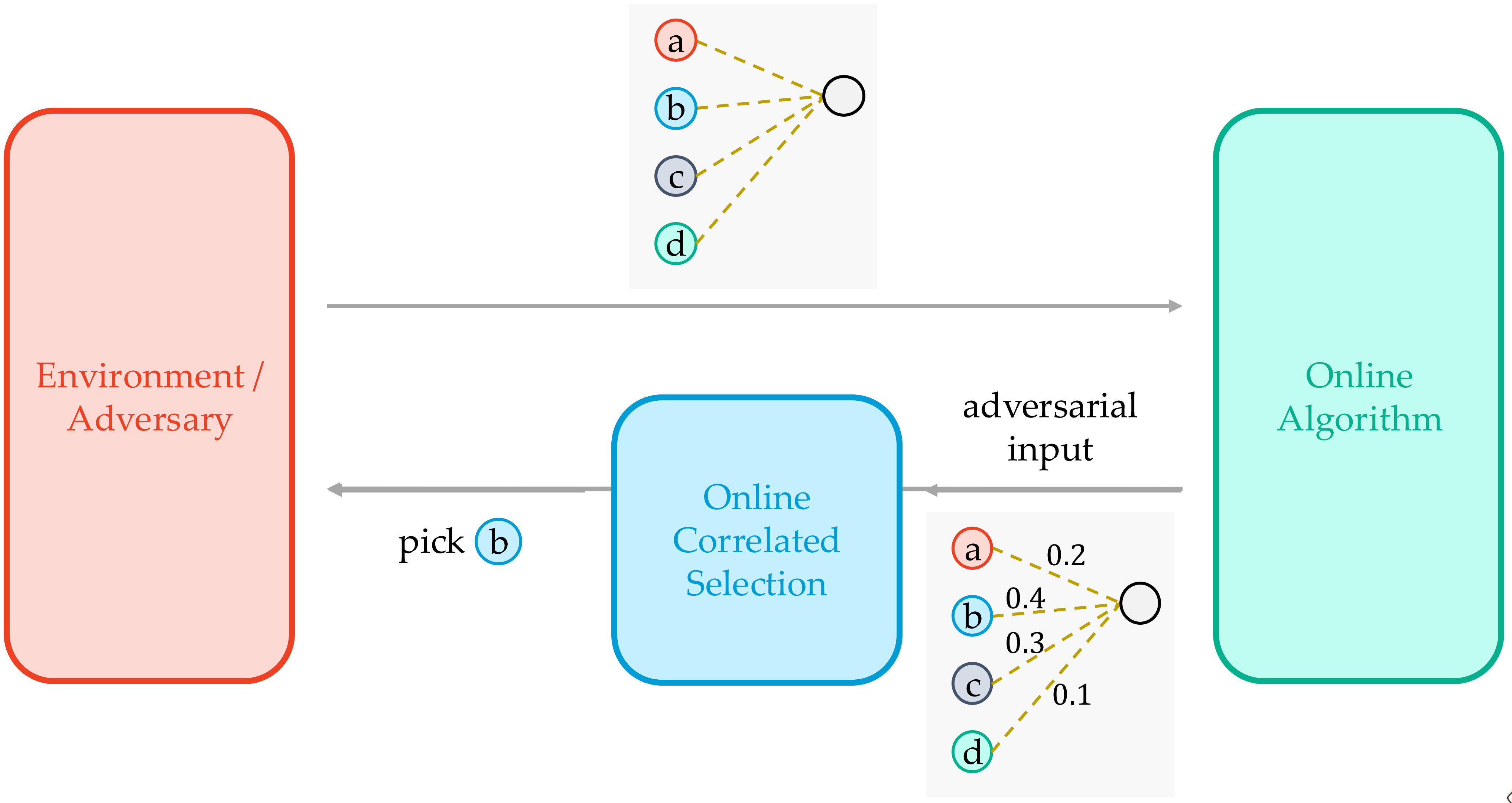
- Rank offline vertices randomly
- Match to the highest-ranked unmatched neighbor

Ranking is $1 - {}^{1}/_{e}$ competitive, which is optimal

From Matching to Submodular Welfare Maximization



Relax-and-Round



Online Correlated Selection (OCS)

- n bins (agent/advertiser/basket)
- At each step t:
  - A ball (item/impression/egg) arrives with $x^t = (x_1^t, x_2^t, \dots, x_n^t)$, where

$$\sum_i x_i^t = 1$$

  - OCS puts the ball into a bin i with a positive x_i^t

Fixed any bin, the randomness is "negatively correlated":
if it is not selected this time, it gets a higher chance next time

Today's Agenda

Part I

- ❑ Intro: What is Online Correlated Selection?
- ❑ Optimal two-way OCS for matching
- ❑ OCS for Display Ads

Part II

- ❑ Multi-way OCS for matching
- ❑ Optimal two-way Stochastic OCS for matching
- ❑ Outro: Summary, directions, and connections to other problems

Two-Way OCS

- Each ball can only go to two bins, with 50-50 fractional allocation

> **Independent Random Selection:** Put each ball into a bin randomly with a fresh random bit.

- If a bin is incident to k balls, it is empty with probability:

$$\frac{1}{2^k}$$

Can we do better?

Optimal Two-Way OCS

Gao, He, H., Nie, Yuan, and Zhong (2021)

- If only one of two bins is empty, put the ball into the empty bin
- What if **both bins are empty**?
  - A fresh coin flip doesn't lead to improvement in the worst case
  - **Test your intuition:** If the bins currently have **degrees** $\boldsymbol{d_1 < d_2}$, shall we put the ball into bin 1 or bin 2?

> **Optimal Algorithm:** Put each ball into the bin with **larger degree**, break ties randomly

- If a bin is incident to k balls, it is empty with probability:

$$\frac{1}{2^{2^k-1}}$$

exponential improvement compared to baseline

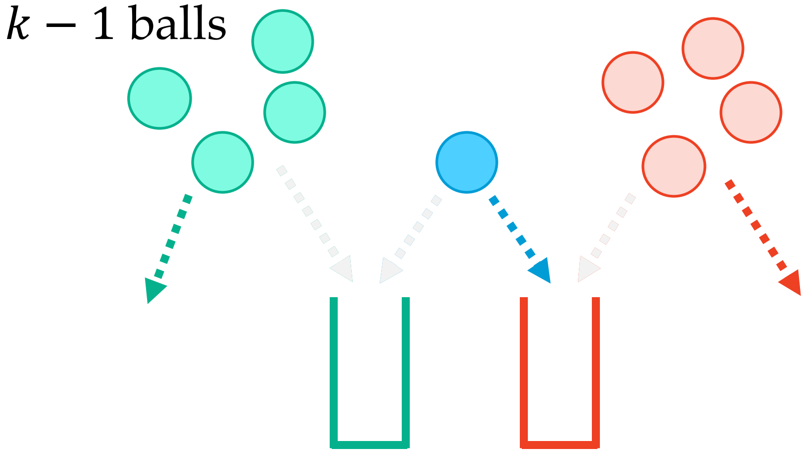
Desired bound: $\frac{1}{2^{2^k-1}}$

Proof by Induction

- Suppose it is true for bins of degree $k-1$
- Consider a bin G and the k^{th} incident (blue) ball
- Consider the other bin R incident to the blue ball

If bin G is empty:

- No green balls go to bin G
- No red balls go to bin R
- Blue ball goes to bin R



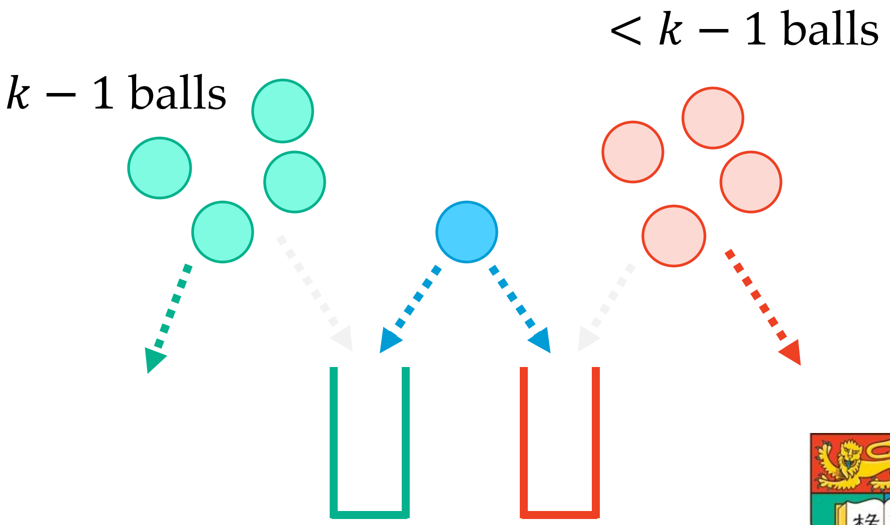
Desired bound: $\frac{1}{2^{2^k-1}}$

Case 1: At most $k-2$ red balls

- Blue ball must go to bin G (if still empty)
- Bin G is empty with probability zero

If bin G is empty:

- No green balls go to bin G
- No red balls go to bin R
- Blue ball goes to bin R



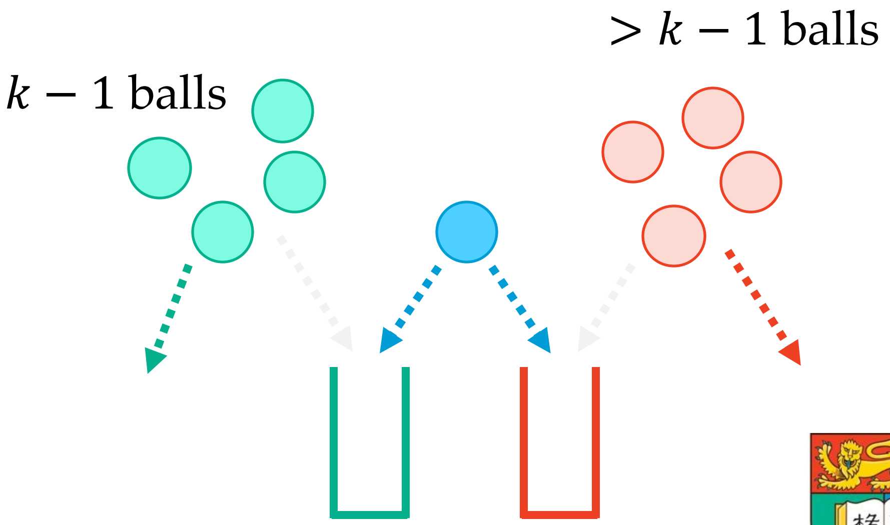
Desired bound: $\frac{1}{2^{2^k-1}}$

Case 2: At least k red balls

- No red balls go to bin R with probability at most $\frac{1}{2^{2^k-1}}$
- Bin G is empty with probability at most $\frac{1}{2^{2^k-1}}$

If bin G is empty:

- No green balls go to bin G
- No red balls go to bin R
- Blue ball goes to bin R



Desired bound: $\frac{1}{2^{2^k-1}}$

Case 3: $k-1$ red balls

$$\mathbf{Pr}[\text{bins } G, R \text{ are both empty after } k-1 \text{ incident balls}] \cdot \frac{1}{2}$$

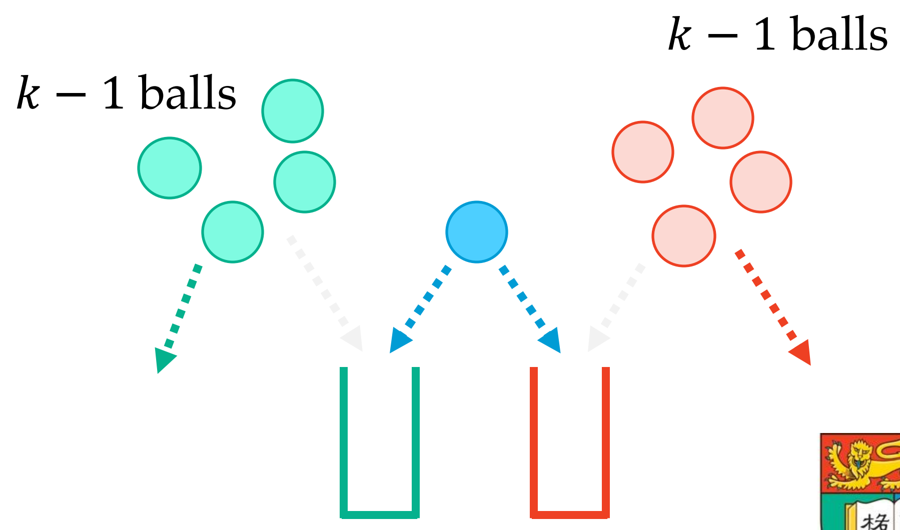
negative correlation only for two-way

$$\leq \mathbf{Pr}[\text{bin } G \text{ empty}] \cdot \mathbf{Pr}[\text{bin } R \text{ empty}] \cdot \frac{1}{2}$$

$$\leq \frac{1}{2^{2^{k-1}-1}} \cdot \frac{1}{2^{2^{k-1}-1}} \cdot \frac{1}{2} = \frac{1}{2^{2^k-1}}$$

If bin G is empty:

- ❑ No green balls go to bin G
- ❑ No red balls go to bin R
- ❑ Blue ball goes to bin R



Takeaway: Optimal Two-Way OCS

Gao, He, H., Nie, Yuan, and Zhong (2021)

Put each ball into an empty bin with **larger degree**, break ties randomly.

- If a bin is incident to k balls, it is empty with probability:

$$\frac{1}{2^{2^k-1}}$$

exponential improvement compared to baseline

- Proof by induction
- Use a negative correlation property only true for two-way
- Optimal 0.536 two-choice algorithm, using only $\log\log n$ random bits
  Gao et al. (2021); Buchbinder, Naor, and Wajc (2023)

Today's Agenda

Part I

- ❑ Intro: What is Online Correlated Selection?
- ❑ Optimal two-way OCS for matching
- ❑ OCS for Display Ads

Part II

- ❑ Multi-way OCS for matching
- ❑ Optimal two-way Stochastic OCS for matching
- ❑ Outro: Summary, directions, and connections to other problems

Display Ads

Feldman, Korula, Mirrokni, Muthukrishnan, and Pál (2009)

Offline vertex can be rematched; only **heaviest edge** counts (i.e., disposing previous lower-weight edges for free)

Until 2020:

- Greedy is 0.5-competitive
- Balance algorithm is $1 - {}^{1}/_{e}$ competitive
- No better integral algorithm was known

Display Ads with OCS

Interpret as a coverage function over $[0, \infty)$

Devanur, H., Korula, Mirrokni, and Yan (2013)

$$\mathbf{E}[\text{heaviest weight}]$$

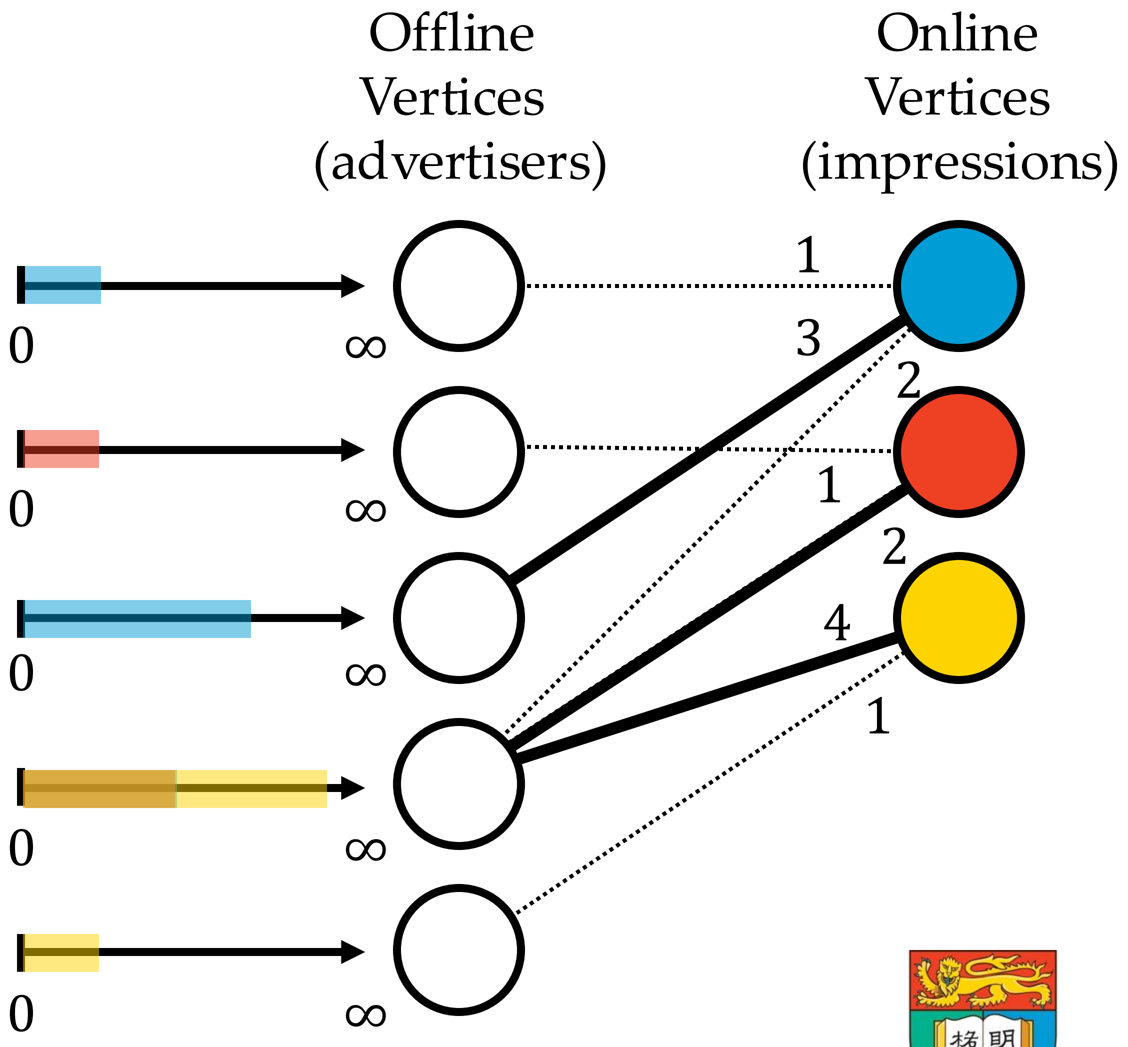
$$= \int_0^\infty \mathbf{Pr}[\text{heaviest weight} \geq w] \, \mathrm{d}w$$

Two-Way γ-OCS for Display Ads

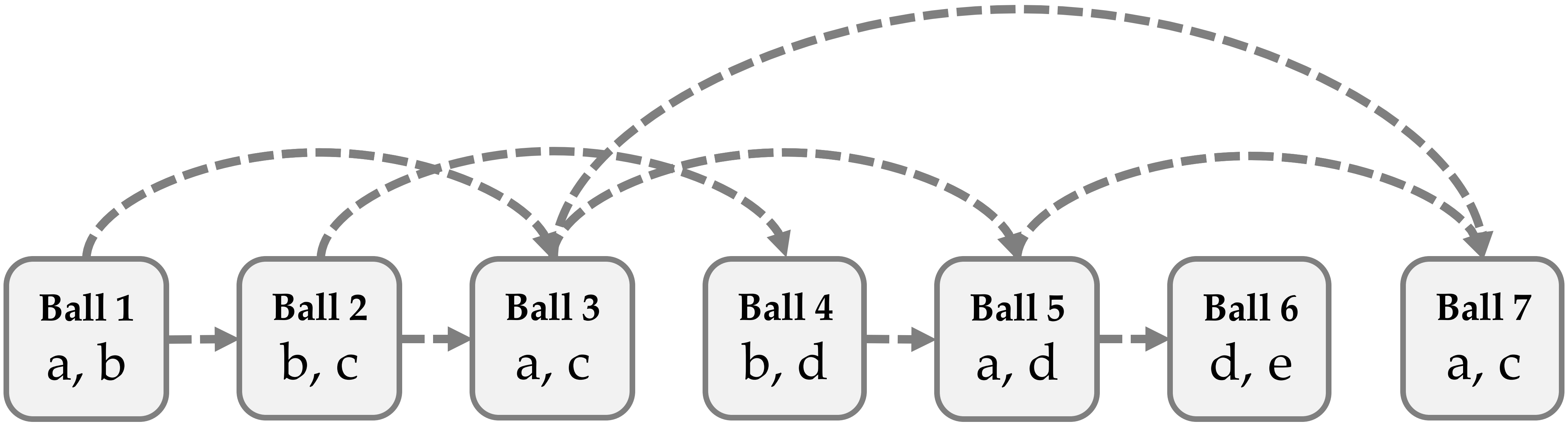
Fahrbach, H., Tao, and Zadimorghaddam (2020)

If a bin is incident to k balls with weight $\geq w$, **with g "gaps",** it gets one of these balls w.p. at least

$$1 - 2^{-k}(1-\gamma)^{k-1-g}$$

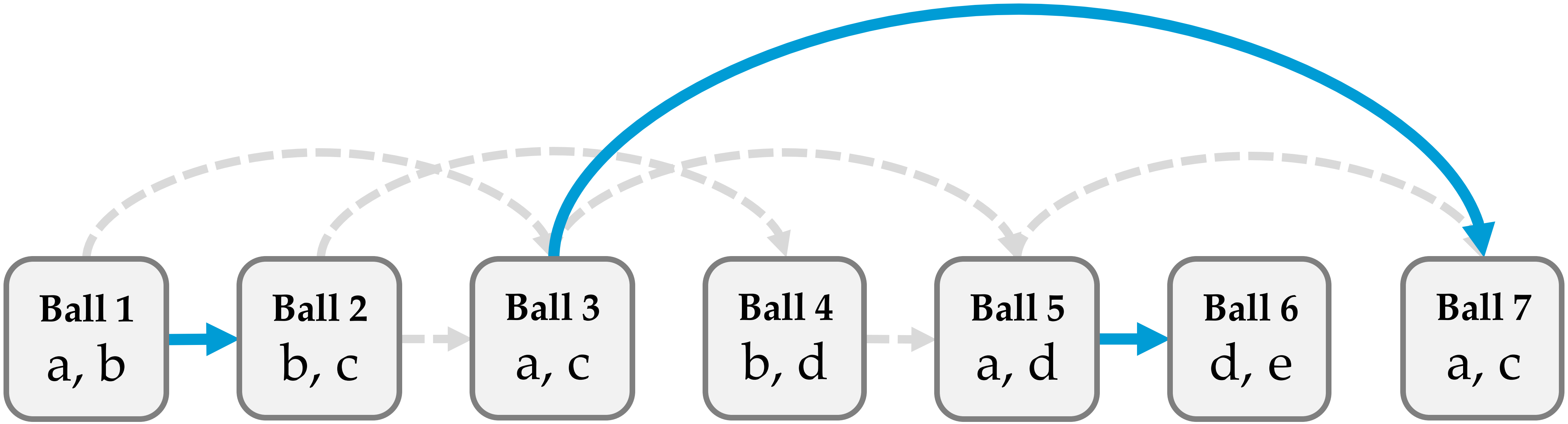


Dependence Graph (Dashed Edges)



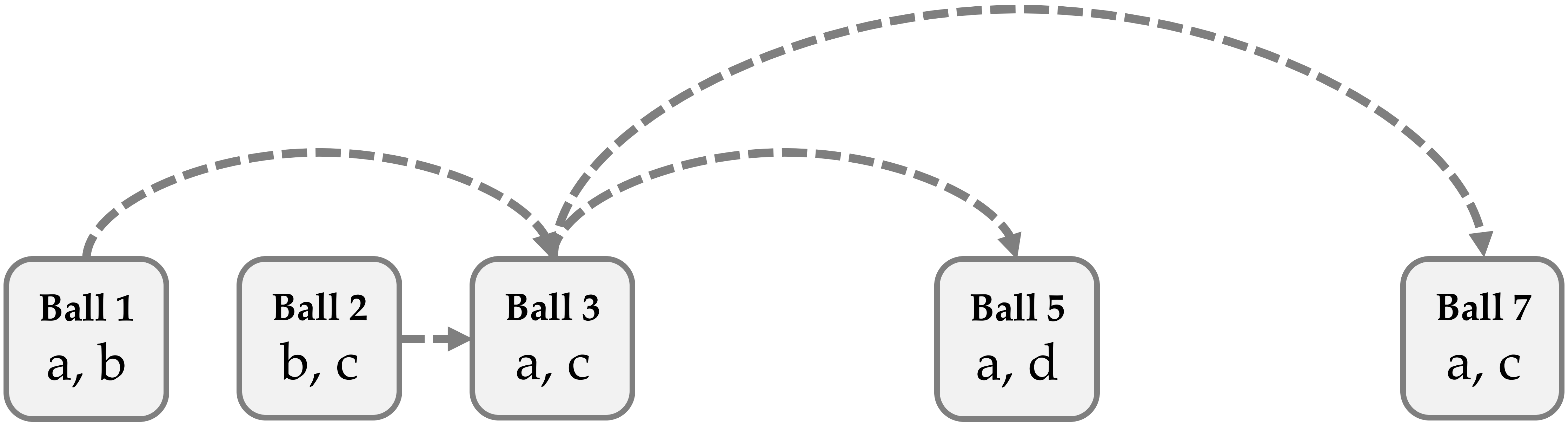
- Let there be a directed path along the balls incident to each bin
- OCS **may** (perfectly) negatively correlate the allocation of two balls connected by an edge

Matching in Dependence Graph (Solid Edges)



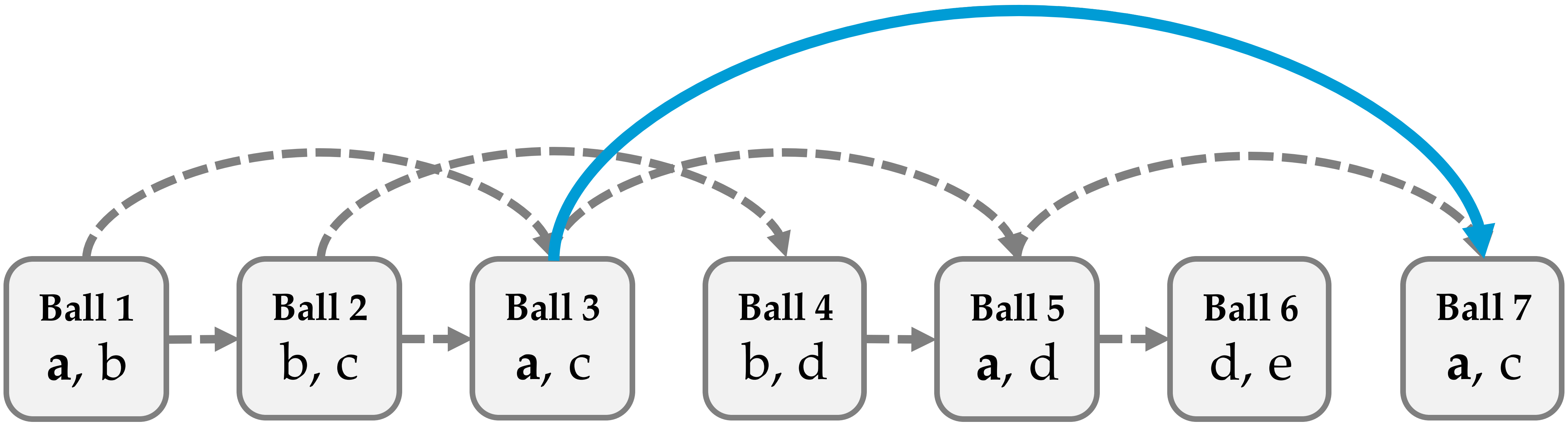
- OCS perfectly negatively correlates the allocation of two matched balls

Finding a (Random) Matching



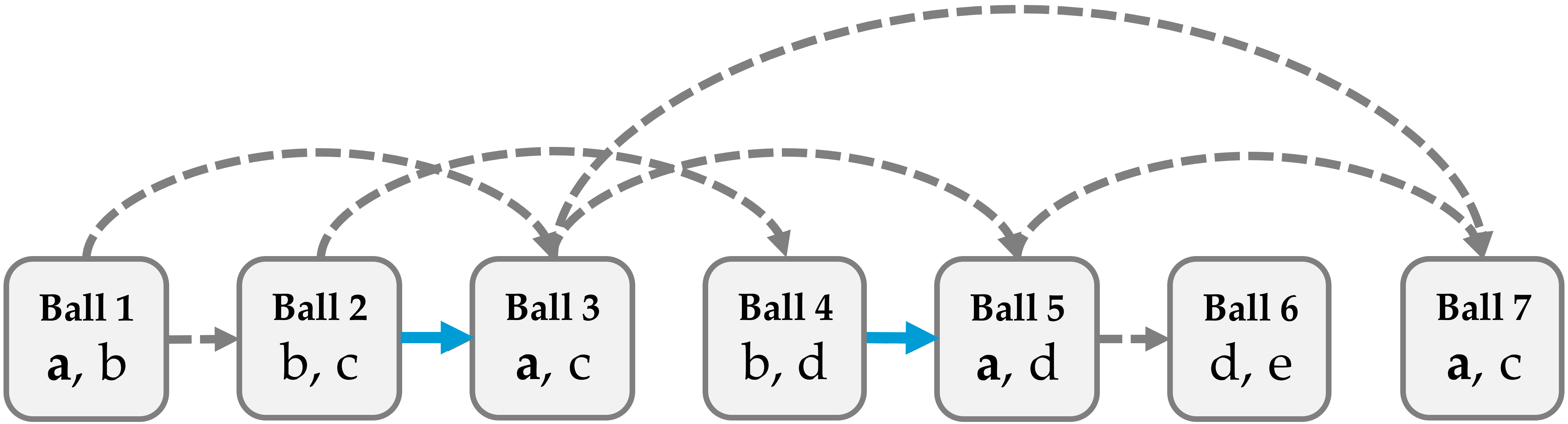
- Each ball randomly picks one of the 4 incident edges
- An edge is in the **matching** if **both** endpoints pick it

Analysis of a 1/16-OCS



- If two balls incident to bin a are matched
  (even if it was due to another bin)
- Bin a gets one of the two balls with certainty

Analysis of a 1/16-OCS (cont'd)



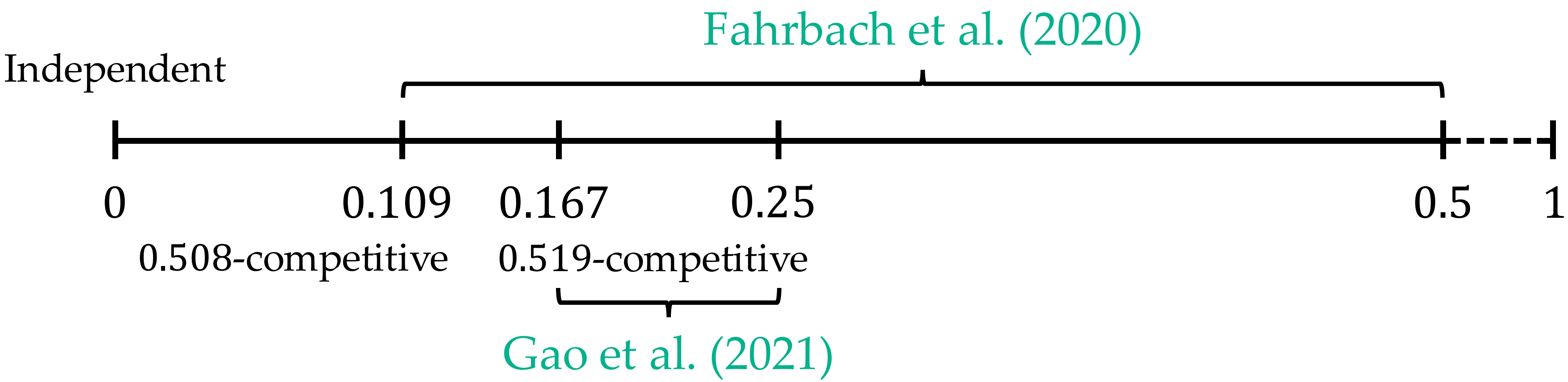
- Otherwise…
- These balls are allocated with independent random bits
  (from bin *a*'s viewpoint)

If a bin doesn't get any of k balls with weight $\geq w$

- These balls are **not** matched
  - There are at least $k - 1 - g$ edges due to this bin
  - Each will be matched with prob. $1/16$
  - This happens with probability $\leq (1 - 1/16)^{k-1-g}$ (negative association)
- All k balls (independently and randomly) go to other bins
  - This happens with probability 2^{-k}

Takeaway: Display Ads with OCS

- Two-Way OCS for Display Ads



- Multi-Way OCS for Display Ads
  - Shin and An (2021): Three-way, 0.509-competitive
  - Blanc and Charikar (2021): General, 0.536-competitive

AdWords

Mehta, Saberi, Vazirani, and Vazirani (2005)

Offline vertex can match many times with **budget-additive** gain

$$\min \left\{ \sum w_{ij} , B_i \right\}$$

Until 2020:

- Greedy is 0.5-competitive
- Balance algorithm is $1 - {}^1/_e$ competitive
- No better integral algorithm was known

With OCS

- 0.504-competitive algorithm for AdWords

  H., Zhang, and Zhang (2020); Chen, H., and Sun (2024)

Today's Agenda

Part I

- ❑ Intro: What is Online Correlated Selection?
- ❑ Optimal two-way OCS for matching
- ❑ OCS for Display Ads

Part II

- ❑ Multi-way OCS for matching
- ❑ Optimal two-way Stochastic OCS for matching
- ❑ Outro: Summary, directions, and connections to other problems

Multi-Way OCS for Matching

Gao, He, H., Nie, Yuan, and Zhong (2021)

At each step t:

- A ball arrives with $x^t = (x_1^t, x_2^t, \dots, x_n^t)$
- OCS puts the ball into a bin i with a positive x_i^t
- Let $y_i = \sum_t x_i^t$ be the total allocation to bin i

Baseline

- Allocate to bin i w.p. x_i^t
- Independently in each step t

$$\Pr[i \text{ is empty}] = \prod_t \left(1 - x_i^t\right) \le e^{-y_i}$$

(Overly-)Idealized Benchmark

- Allocate to bin i w.p. x_i^t
- Somehow, no bin gets ≥ 2 balls

$$\Pr[i \text{ is empty}] = 1 - y_i = e^{-y_i - \frac{1}{2} y_i^2 - \frac{1}{3} y_i^3 - \cdots}$$

Gao et al. (2021)

The following algorithm ensures that

$$\Pr[i \text{ is empty}] \leq p(y_i) = e^{-y_i - \frac{1}{2} y_i^2 - \frac{4-2\sqrt{3}}{3} y_i^3}$$

At step t, allocate to an **empty bin** i w.p. proportional to

$$\frac{x_i^t}{p\left(y_i^{t-1}\right)}$$

Baseline

- Allocate to bin i w.p. x_i^t
- Independently in each step t

$$\Pr[i \text{ is empty}] = \prod_t \left(1 - x_i^t\right) \leq e^{-y_i}$$

(Overly-)Idealized Benchmark

- Allocate to bin i w.p. x_i^t
- Somehow, no bin gets ≥ 2 balls

$$\Pr[i \text{ is empty}] = 1 - y_i = e^{-y_i - \frac{1}{2} y_i^2 - \frac{1}{3} y_i^3 - \cdots}$$

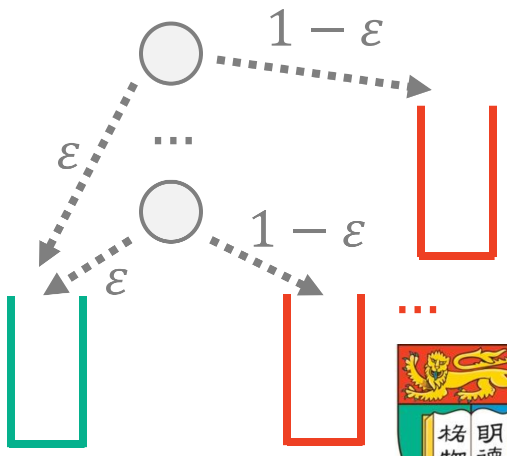
Intuition

Improved Baseline?

- Allocate to an **empty bin** i w.p. proportional to x_i^t
- Independently in each step t

- This still only gives e^{-y_i} in the worst case
- **Idea:** Favor bins with larger y_i?
  - Red bin has extra chance if green bin got a ball in the past
  - Red bin can afford to "return the favor" to an **empty** green bin
- With our scaling, bin i's expected sampling mass is

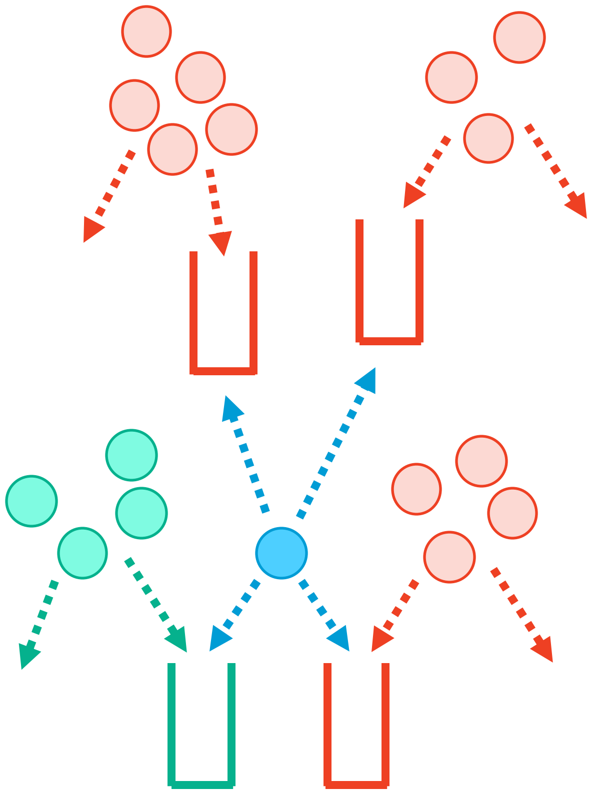
$$\mathbf{E}\, \frac{x_i^t}{p(y_i^{t-1})} \cdot \mathbf{1}_{i \text{ is empty}}^{t-1} \leq x_i^t$$



let $w_i^{t-1} = \frac{1}{p(y_i^{t-1})}$ for ease of notations

"Proof" by Induction

- Suppose it is true for time $t-1$
- Consider bin i incident to the (blue) ball at time t
- Consider the other bins incident to the blue ball



$\mathbf{Pr}[\text{bin } i \text{ is empty}]$

negative correlation may not hold

$$\overset{?}{\leq} p(y_i^{t-1}) \cdot \mathbf{E} \frac{\sum_{j \neq i} w_j^{t-1} \cdot x_j^t \cdot \mathbf{1}_{j\text{ empty}}^{t-1}}{w_i^{t-1} \cdot x_i^t + \sum_{j \neq i} w_j^{t-1} \cdot x_j^t \cdot \mathbf{1}_{j\text{ empty}}^{t-1}}$$

$$\leq p(y_i^{t-1}) \frac{\sum_{j \neq i} x_j^t}{w_i^{t-1} \cdot x_i^t + \sum_{j \neq i} x_j^t}$$

$$= p(y_i^{t-1}) \frac{1 - x_i^t}{w_i^{t-1} \cdot x_i^t + 1 - x_i^t} \leq p(y_i^{t-1} + x_i^t)$$

this recurrence determines $p(y)$

let $w_i^{t-1} = \frac{1}{p(y_i^{t-1})}$ for ease of notations

Lifted Induction on Joint Events

$$\mathbf{Pr}[j \text{ empty} \mid i \text{ empty}] = \frac{\Pr[i, j \text{ both empty}]}{\Pr[i \text{ empty}]}$$

induction hypothesis does not apply without negative correlation

induction hypothesis applies

Lifted Claim. For any subset of bins S, $\mathbf{Pr}[S \text{ is empty}] \le \prod_{i \in S} p(y_i)$.

$$\mathbf{Pr}[i \text{ empty}] = \mathbf{E}\left[\mathbf{1}_{i \text{ empty}}^{t-1} \cdot \frac{w_j^{t-1} \cdot x_j^t \cdot \mathbf{1}_{i,j \text{ empty}}^{t-1}}{w_i^{t-1} \cdot x_i^t \cdot \mathbf{1}_{i \text{ empty}}^{t-1} + w_j^{t-1} \cdot x_j^t \cdot \mathbf{1}_{i,j \text{ empty}}^{t-1}}\right]$$

$$\le p(y_i^{t-1}) \cdot \frac{x_j^t \, p(y_i^{t-1})}{x_i^t + x_j^t \, p(y_i^{t-1})}$$

By Jensen, concavity, and monotonicity of $\frac{xy}{x+y}$

Applications of Multi-Way OCS for Matching

- Rounding **Balance*** with **Multi-Way OCS** is 0.593 competitive for (unweighted/vertex-weighted) matching

  Gao et al. (2021)

- Approximate class fairness in online allocation

  Hosseini, H., Igarashi, and Shah (2023)

- Breaking $1 - \frac{1}{e}$ barrier Non-IID Online Stochastic Matching

  Tang, Wu, and Wu (2022)

Today's Agenda

Part I

- ❑ Intro: What is Online Correlated Selection?
- ❑ Optimal two-way OCS for matching
- ❑ OCS for Display Ads

Part II

- ❑ Multi-way OCS for matching
- ❑ Optimal two-way Stochastic OCS for matching
- ❑ Outro: Summary, directions, and connections to other problems

Online Stochastic Matching

Feldman, Mehta, Mirrokni, and Muthukrishnan (2009)

Bipartite type graph

$$G = (U, V, E)$$

- Offline vertices U, online types V, edges E are known from the beginning
- Each type arrives by a Poisson process (e.g., unit arrival rate)
- Decide how to match each online vertex on its arrival

Algorithms for Online Stochastic Matching

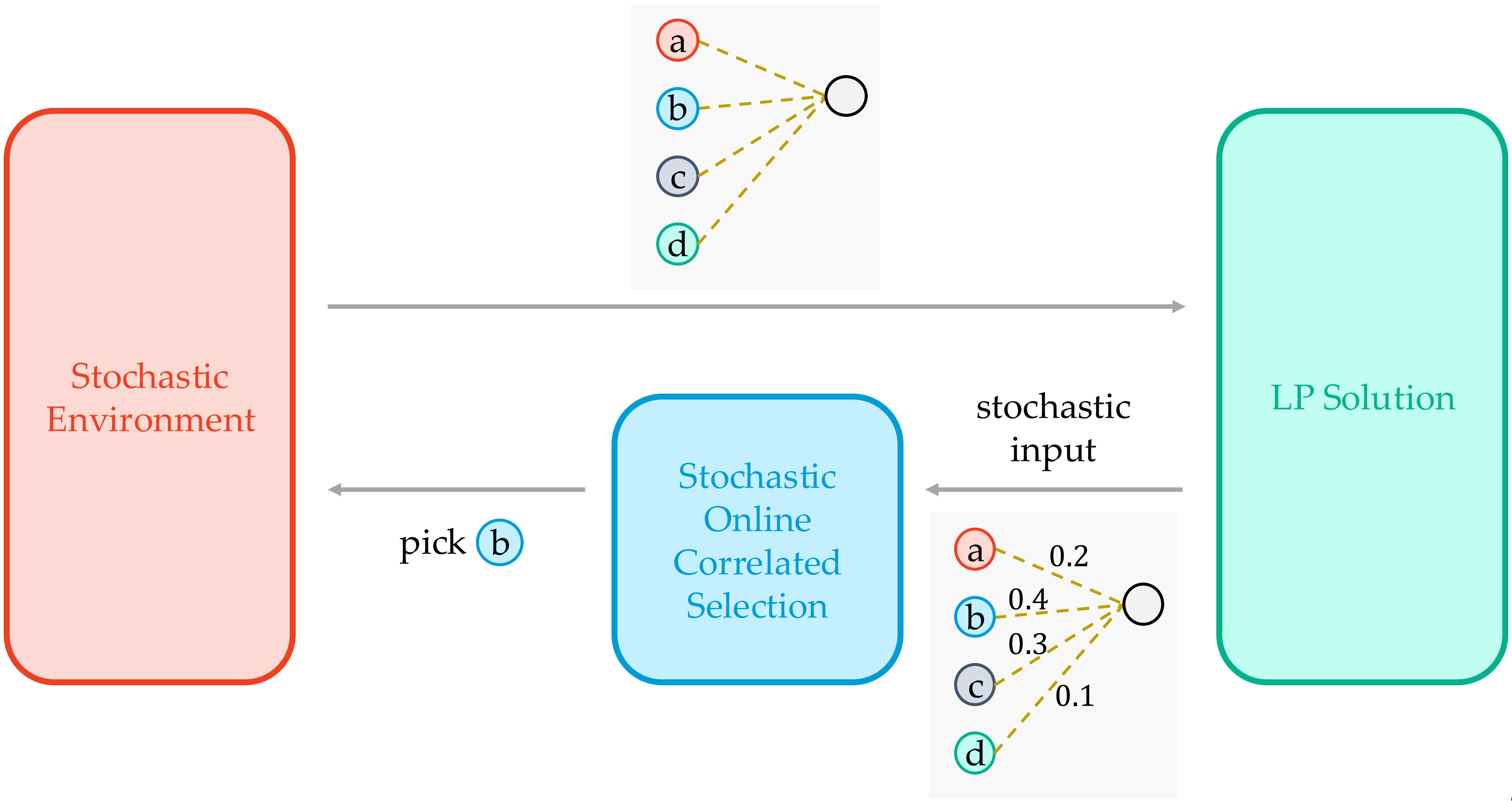
- **LP Relaxation:** Let $x_{ij} \geq 0$ be match prob. of edge $(i, j) \in E$
  Torrico, Ahmed, and Toriello (2018); Huang and Shu (2021)

$$
\begin{aligned}
& \max && \textstyle\sum_{(i,j)\in E} x_{ij} \\
& \text{for any online } j \in V && \textstyle\sum_i x_{ij} \leq 1 \\
& \text{for any offline } i \in U,\ T \subseteq V && \textstyle\sum_{j\in T} x_{ij} \leq 1 - e^{-|T|}
\end{aligned}
$$

- LP solution serves two roles
  - Benchmark for competitive analysis
  - Reference for online matching decisions

online rounding

Relax-and-Round for Online Stochastic Matching

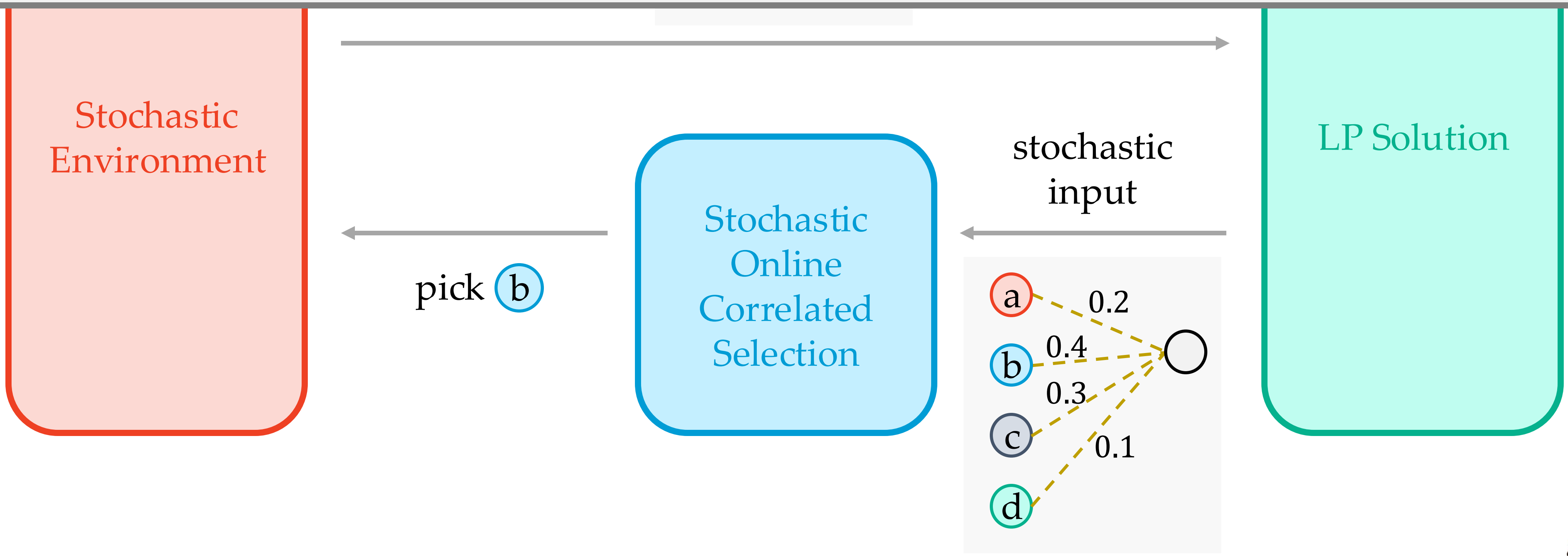


H., Shu, and Yan (2022)

- At step t, allocate to an **empty bin** i w.p. proportional to

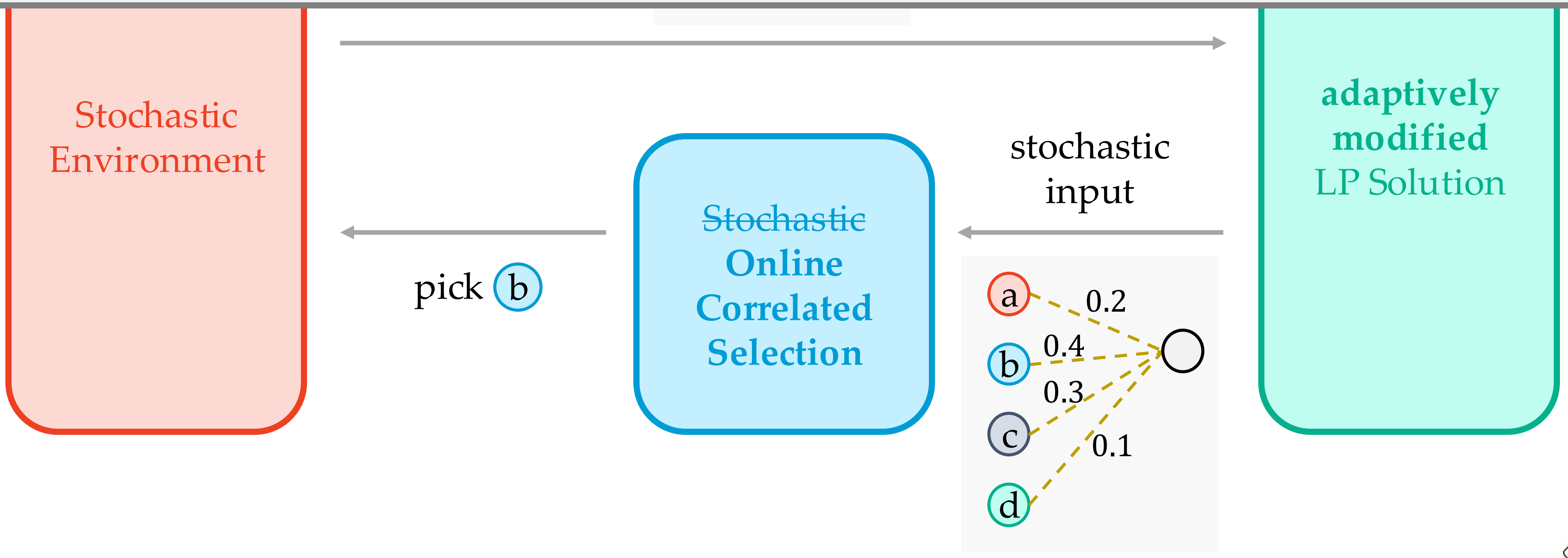
$$x_{ij} \cdot \boxed{e^{t \sum_k x_{ik}} = e^{\mathbf{E}\, y_i}}$$

- 0.716 competitive algorithm for matching



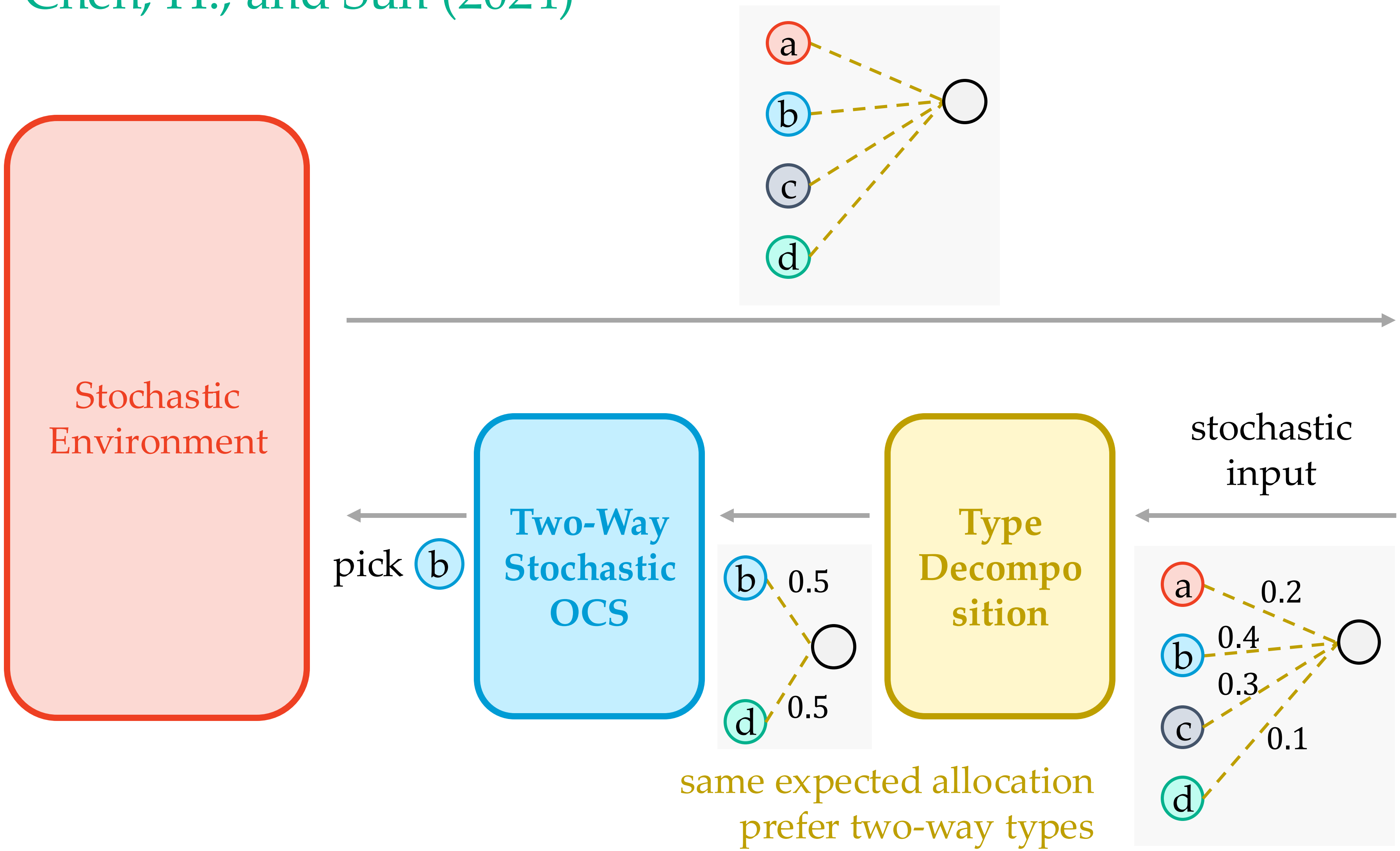
Tang, Wu, and Wu (2022)

- Directly use OCS by Gao et al. (2021)
- Adaptively modify LP solution to reduce variance
- 0.666 competitive algorithm for non-IID model, breaking $1 - {}^{1}/_{e}$ barrier



Reducing Multi-Way to Two-Way

Chen, H., and Sun (2024)



Stochastic Two-Way OCS

- Each type of ball (defined by two bins i, j it can go to) arrives by a Poisson process with arrival rate λ_{ij} in time horizon $[0, 1]$
- Bin i's normalized degree is half of the incident balls' total arrival rate

$$d_i = \frac{1}{2} \sum_{j \neq i} \lambda_{ij}$$

Independent Random Selection: Put each ball into a bin randomly with a fresh random bit.

- Bin i is empty with probability:

$$e^{-d_i}$$

Can we do better?

Optimal Stochastic (Half-Integral) OCS

Chen, H., Sun (2024)

- If only one of two bins is empty, put the ball into the empty bin
- What if both bins are empty?

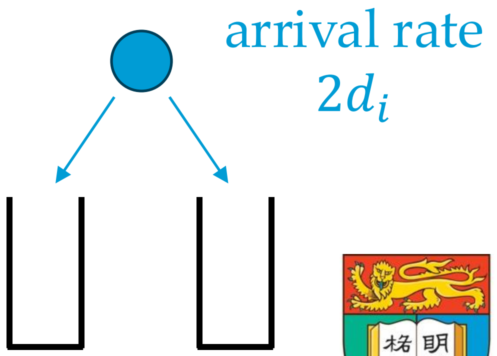
> **Optimal Algorithm:** If a ball arrives at time t, put it into an incident empty bin i with probability proportional to $\boldsymbol{e^{2td_i}}$

- Bin i is empty with probability

$$e^{-2d_i}(1 + d_i)$$

vs. baseline e^{-d_i}

- Tight for two bins and a single type of balls
- Generalize to non-homogenous Poisson arrivals



Proof by "Induction"

Let $p_i(t)$ be the probability that bin i is empty at time t.

$$\frac{\mathrm{d}}{\mathrm{d}t} p_i(t) = \sum_j \lambda_{ij} \left(\mathbf{Pr}[i, j \text{ empty at } t] \cdot \frac{e^{2td_j}}{e^{2td_i} + e^{2td_j}} - p_i(t) \right)$$

$$\leq \sum_j \lambda_{ij} \left(e^{-td_i - td_j} \cdot \frac{e^{2td_j}}{2e^{td_i + td_j}} - p_i(t) \right)$$

$$= \sum_j \lambda_{ij} \left(\frac{1}{2} e^{-2td_i} - p_i(t) \right) = d_i \left(e^{-2td_i} - 2p_i(t) \right)$$

This holds with equality for $p_i(t) = e^{-2td_i}(1 + td_i)$.

Applications of Two-Way Stochastic OCS

Chen, H., and Sun (2024)

- Unweighted and Vertex-Weighted Matching
  - 0.69 competitive algorithm for non-IID model
  - 0.705 competitive algorithm for query-commit model
- Display Ads (a.k.a., edge-weighted matching)
  - 0.644 competitive algorithm for non-IID model, breaking $1 - {}^{1}/_{e}$ barrier
- AdWords
  - 0.633 competitive algorithm for non-IID model, breaking $1 - {}^{1}/_{e}$ barrier
  - IID special case is an open problem from Mehta (2013)

Today's Agenda

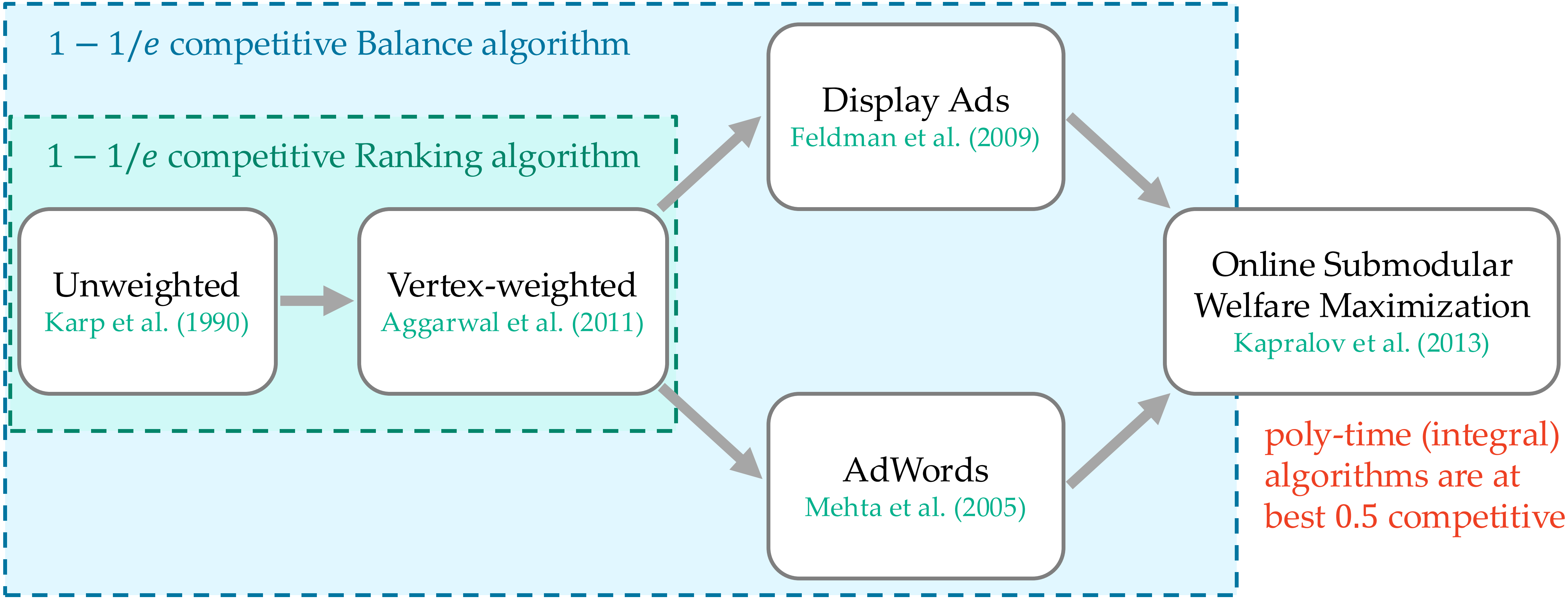
Part I

- ❑ Intro: What is Online Correlated Selection?
- ❑ Optimal two-way OCS for matching
- ❑ OCS for Display Ads

Part II

- ❑ Multi-way OCS for matching
- ❑ Optimal two-way Stochastic OCS for matching
- ❑ Outro: Summary, directions, and connections to other problems

OCS is Online Rounding for Matching and Beyond



Fractional Allocation + Online Rounding works

Direction I: Improve OCS Algorithms

- Two-way stochastic OCS for AdWords

When an online vertex with two-way surrogate type $i \sim \{j, k\}$ arrives at time step t:

1. Pick $m \in \{j, k\}$ uniformly at random and mark this online vertex with m.
2. If this is the second online vertex marked with m, then make the opposite selection to the first one (w.r.t. m).
3. Otherwise, select j or k uniformly at random.

- Two-way stochastic OCS for Display Ads

Direction II: Online Submodular Welfare Maximization

- Poly-time algorithms are at best
  - 0.5 competitive in adversarial model
  - $1 - {}^{1}/_{e}$ competitive in stochastic model
  - even for coverage functions

  Kapralov, Post, and Vondrák (2013)
- Rely on hardness of max coverage, **no information theoretic hardness**
- **$1 - {}^{1}/_{e}$ competitive fractional algorithm** for coverage functions

> Is there an OCS for coverage (even submodular) functions?

Direction III:
Connection to Balls into Bins and Scheduling

- Each type of ball (defined by two bins i, j it can go to) arrives by a Poisson process with arrival rate λ_{ij} in time horizon $[0, 1]$
- Bin i's normalized degree is the incident balls' total arrival rate

$$d_i = \frac{1}{2} \sum_{j \neq i} \lambda_{ij}$$

- Let n_i be the number of balls in bin i at the end

Matching

$$\max \sum_i \mathbf{1}_{n_i \geq 1}$$

Balls into Bins

$$\min \max_i n_i$$

$$\approx \min \sum_i e^{\lambda n_i}$$

Completion Time

$$\min \sum_i \frac{1}{2}\left(n_i^2 + n_i\right)$$

References

- Slides will be on my homepage
- INFORMS'24 Tutorial on Sequential Randomized Rounding
- FOCS'23 Workshop on Online Algorithms and Online Rounding
- WINE'23 Tutorial on Online Matching
- Brief Survey on Online Matching in SIGecom Exchange

Thank you! Questions?