Semi-Online Bipartite Matching

Zoya Svitkina

with Ravi Kumar, Manish Purohit, Aaron Schild, Erik Vee

Google

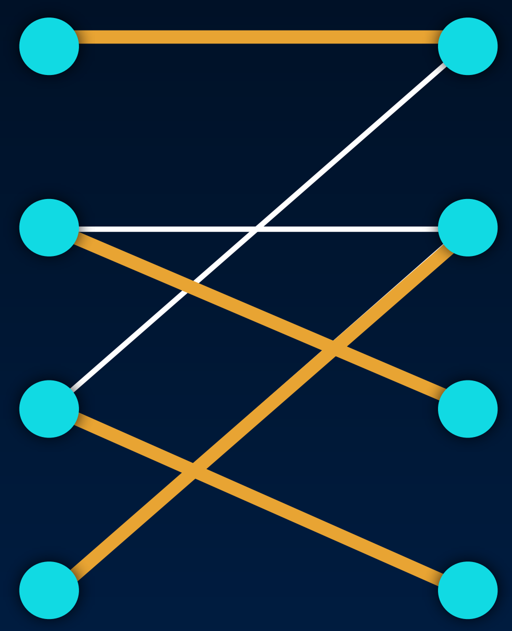
TTIC Summer Workshop on Learning-Based Algorithms
August 13, 2019

Semi-online algorithms

- Future is partly known, partly adversarial
- Pre-process the known part
- Then make irrevokable decisions at each step
- Interpolates between offline and online models

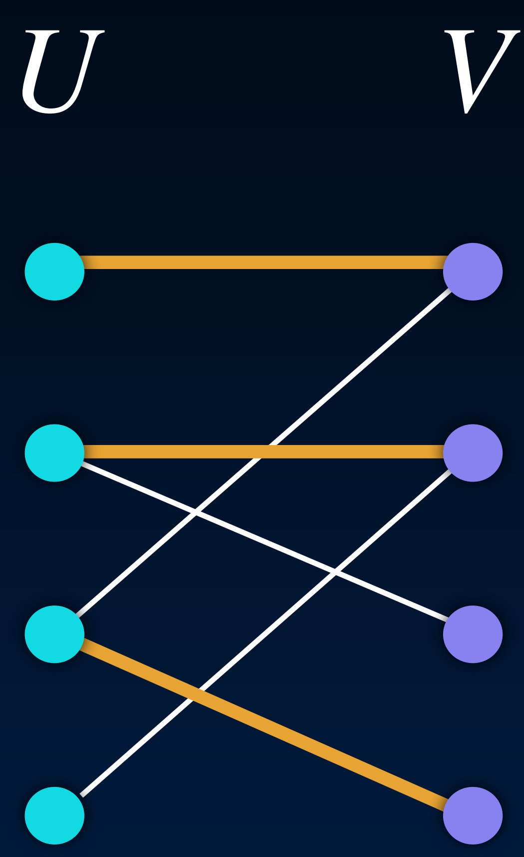
Offline bipartite matching

- Polynomial-time solvable using max flow



Online bipartite matching

- Nodes in U known in advance
- Nodes in V arrive one by one
- Match at each step
- Competitive ratio compares to offline OPT



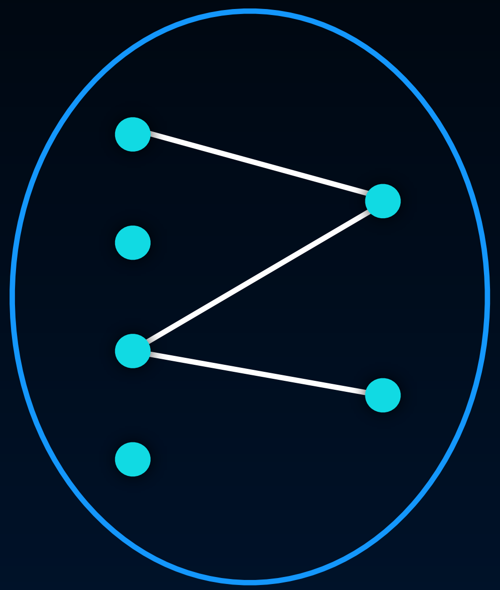
Online bipartite matching

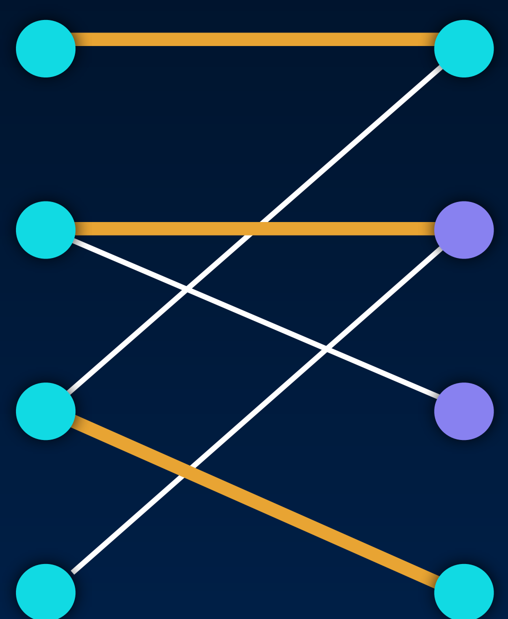
- RANKING algorithm [1] is $1 - 1/e$ competitive:
  - Fix a random permutation of offline nodes
  - For each online node:
    - Match to the first available neighbor in the permutation

[1] Richard Karp, Umesh Vazirani, Vijay Vazirani. An optimal algorithm for on-line bipartite matching. STOC 1990

Semi-online bipartite matching

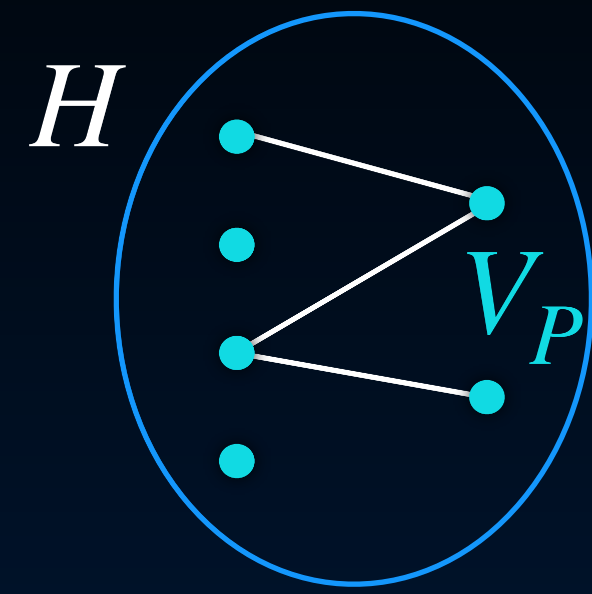
- Know U and part of V in advance
- All of V arrives one by one in arbitrary order
- Match at each step
- Competitive ratio compares to offline OPT
- Integral or fractional matching

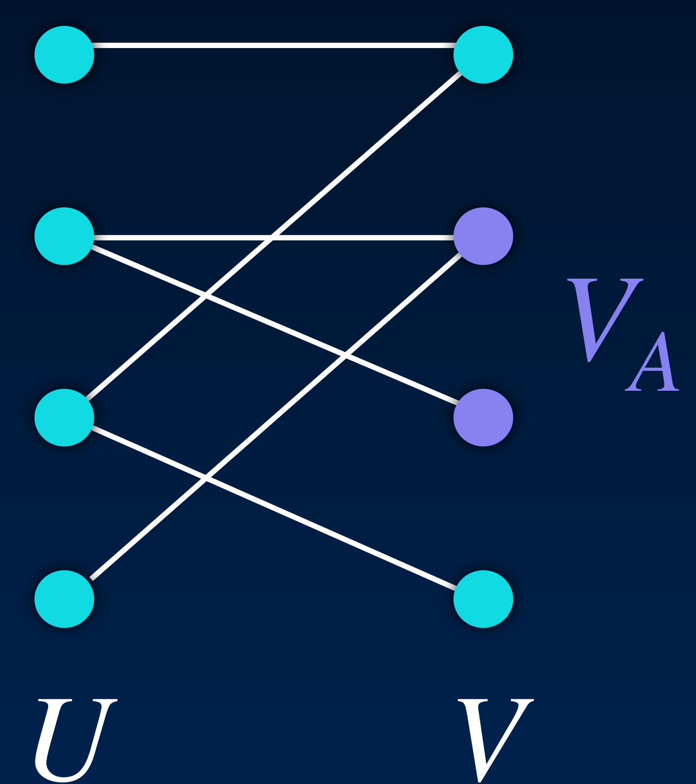




Notation

- Bipartite graph $G = (U, V, E_G)$
- $V = V_P \cup V_A$
  - V_P: known (predicted) part of V
  - V_A: unknown (adversarial) part of V
- Known subgraph $H = (U, V_P, E_H)$





Online/offline parameter δ

- Simplifying assumption for this talk: perfect matching in G
  - $\delta = \frac{|V_A|}{|V|}$, fraction of adversarial nodes
    - $\delta = 0$: offline, $\delta = 1$: online
  - Competitive ratio in terms of δ
- General case: $\delta = 1 - \frac{OPT(H)}{OPT(G)}$
  - Other definition doesn't work if many isolated nodes

Results

- Integral matching:
  - Algorithm with competitive ratio $1 - \delta + \delta^2(1 - 1/e)$
  - Hardness of $1 - \delta e^{-\delta}$ $(\approx 1 - \delta + \delta^2 - \delta^3/2 + \ldots)$
- Fractional matching:
  - Algorithm and hardness of $1 - \delta e^{-\delta}$

Related settings

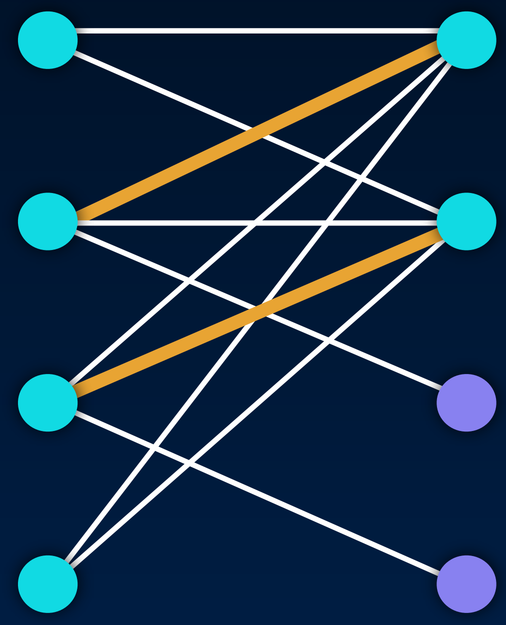
- Optimal online assignment with forecasts
  Erik Vee, Sergei Vassilvitskii, and Jayavel Shanmugasundaram. EC 2010
    - Uncertainty in demands, not in graph structure
- Online allocation with traffic spikes: Mixing adversarial and stochastic models
  Hossein Esfandiari, Nitish Korula, and Vahab Mirrokni. EC 2015
    - Forecast is a distribution, not a fixed graph
    - Large degree assumption
    - Same hardness result
- Maximum matching in the online batch-arrival model
  Euiwoong Lee and Sahil Singla. IPCO 2017
    - Online nodes arrive in batches

Observations

- Worst case: predicted nodes before adversarial
    - Algorithm for this case can be transformed into one for arbitrary order
- Should select a maximum matching on H
    - No benefit to leaving predicted nodes unmatched
    - Do this as preprocessing

Selecting a matching for H

- Any deterministic algorithm would do badly

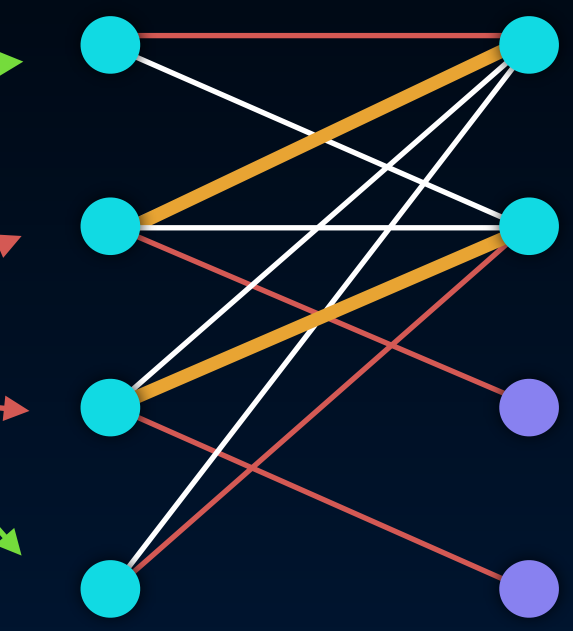


Algorithm outline

- Find a (randomized) maximum matching in H
  - Which nodes to "reserve" for V_A?
- Run RANKING for adversarial nodes

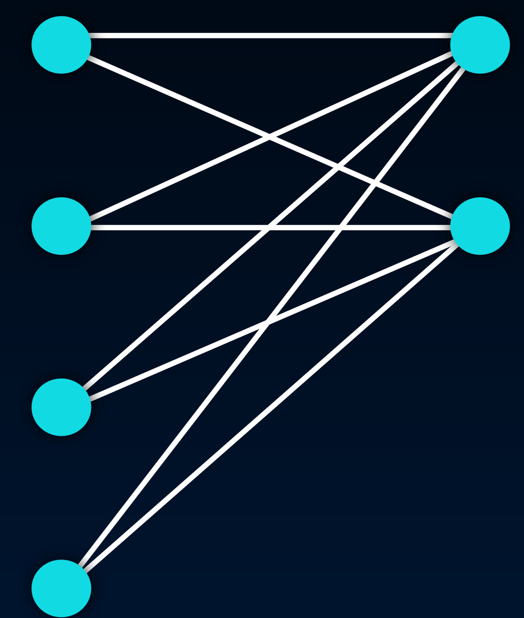
Analysis outline

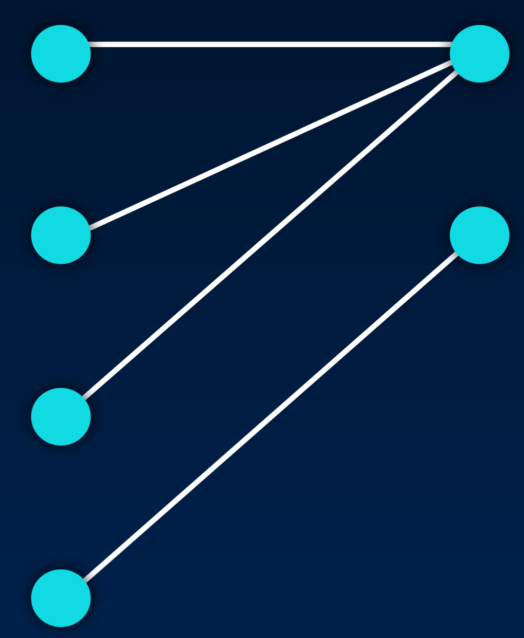
- $Reserved \subseteq U$: not matched in H. $|Reserved| = n - |V_P| = \delta n$
- $Marked \subseteq U$: matched to V_A by OPT. $|Marked| = |V_A| = \delta n$
- Suppose $\mathbb{E}[|Reserved \cap Marked|] = x \cdot n$
  - Matching size $n - \delta n + (1 - 1/e)xn$
  - Competitive ratio $1 - \delta + (1 - 1/e)x$
- Aim for $x = \delta^2$



Reserving nodes

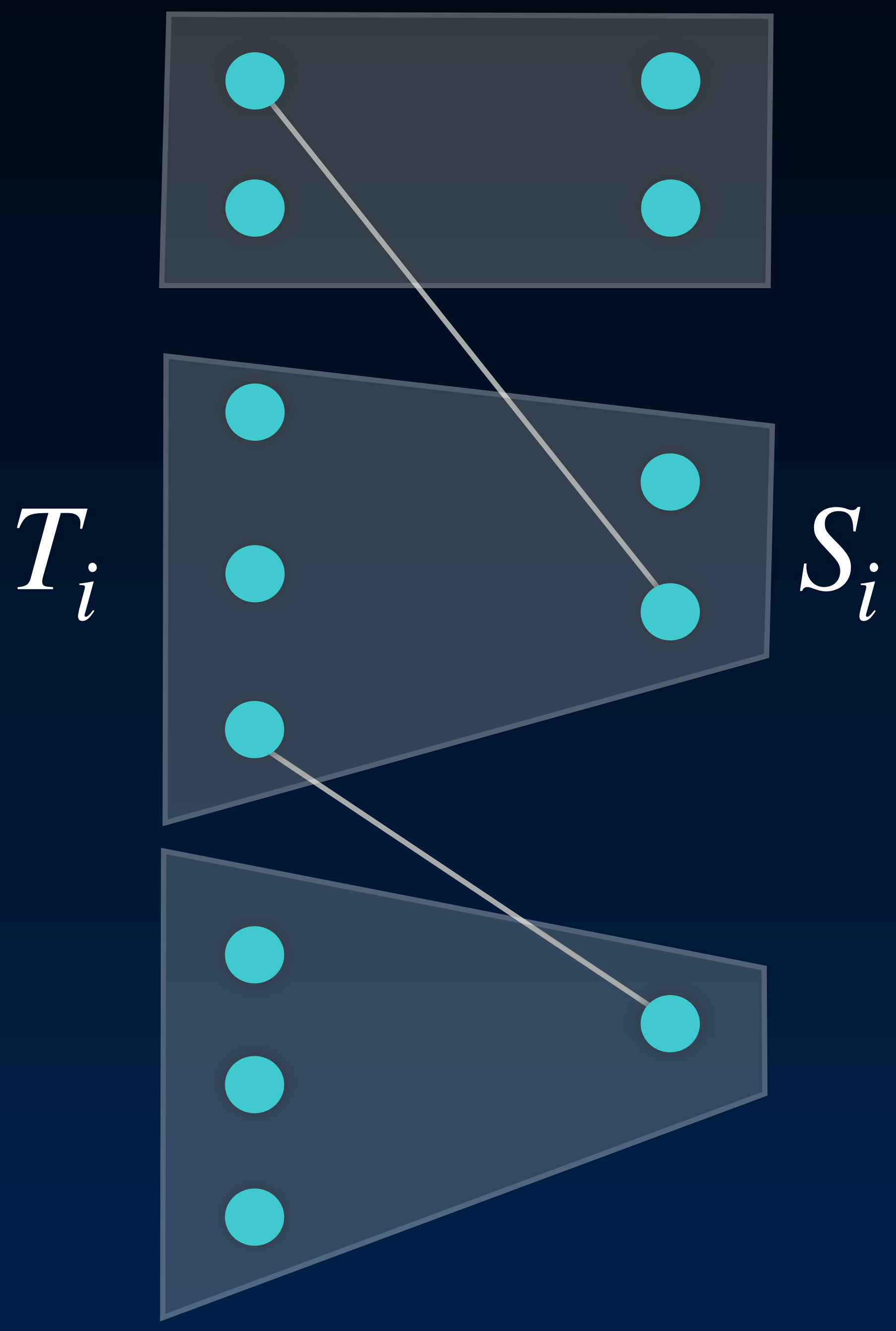
- Goal: sample a matching in H s.t. $\mathbb{E}[\,|Reserved \cap Marked|\,] = \delta^2 n$
- Special case: H is complete
  - Reserve each node with probability δ
- In general, a distribution over matchings s.t. $\forall u \in U,\ \Pr[u \text{ is reserved}] = \delta$ may not exist
- Want a distribution making nodes' probabilities of being reserved as equal as possible





Matching skeleton decomposition[1]

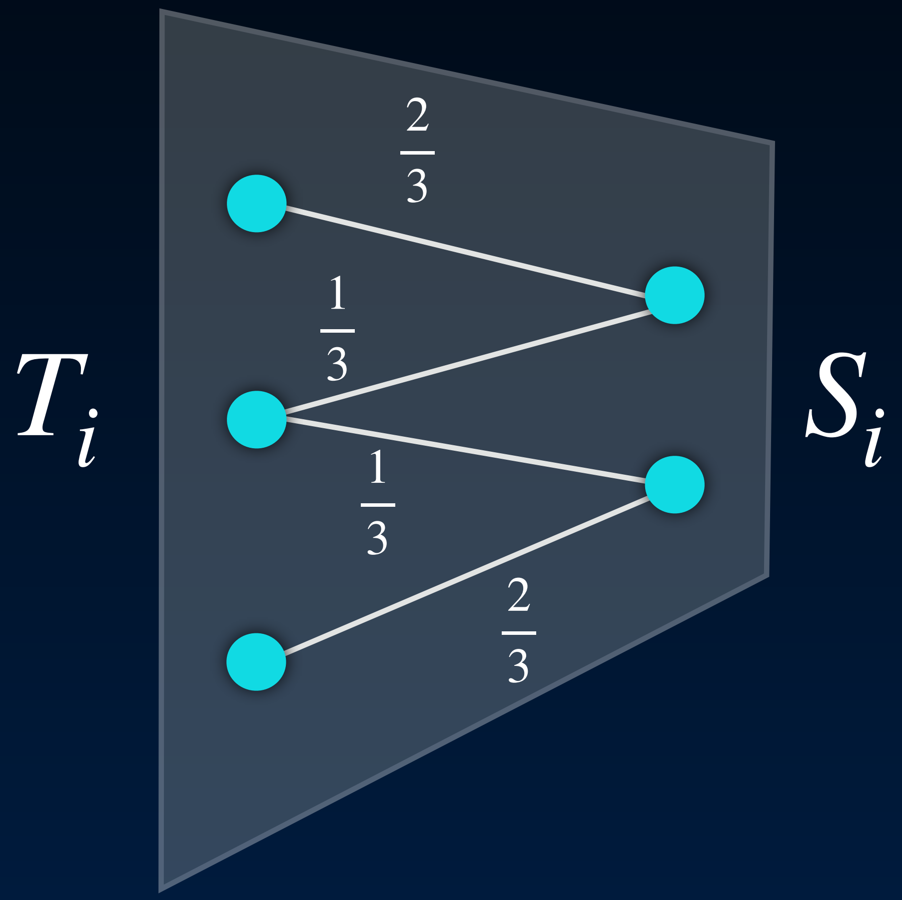
- Decomposition of H (poly-time)
  - $U = \cup_i T_i, \quad V_P = \cup_i S_i$
  - $\Gamma(\cup_{i<j} S_i) = \cup_{i<j} T_i$
  - $i < j \Rightarrow \frac{S_i}{T_i} > \frac{S_j}{T_j}$
  - Fractional matching in each component
    - $\deg(u) = 1, \quad \deg(v) = |S_i| / |T_i|$



[1] Ashish Goel, Michael Kapralov, Sanjeev Khanna.
On the communication and streaming complexity of maximum bipartite matching. SODA 2012.

Dependent rounding

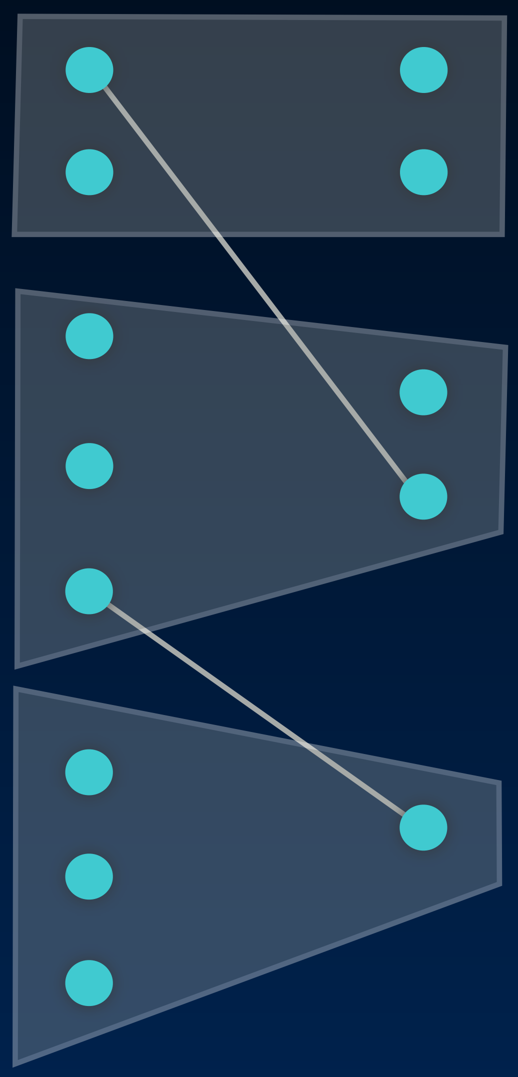
- Apply dependent rounding [1] to each component of the matching skeleton
  - Let $d_i = |T_i| - |S_i|$
  - Probability of $u \in T_i$ being reserved is $\dfrac{d_i}{|T_i|}$



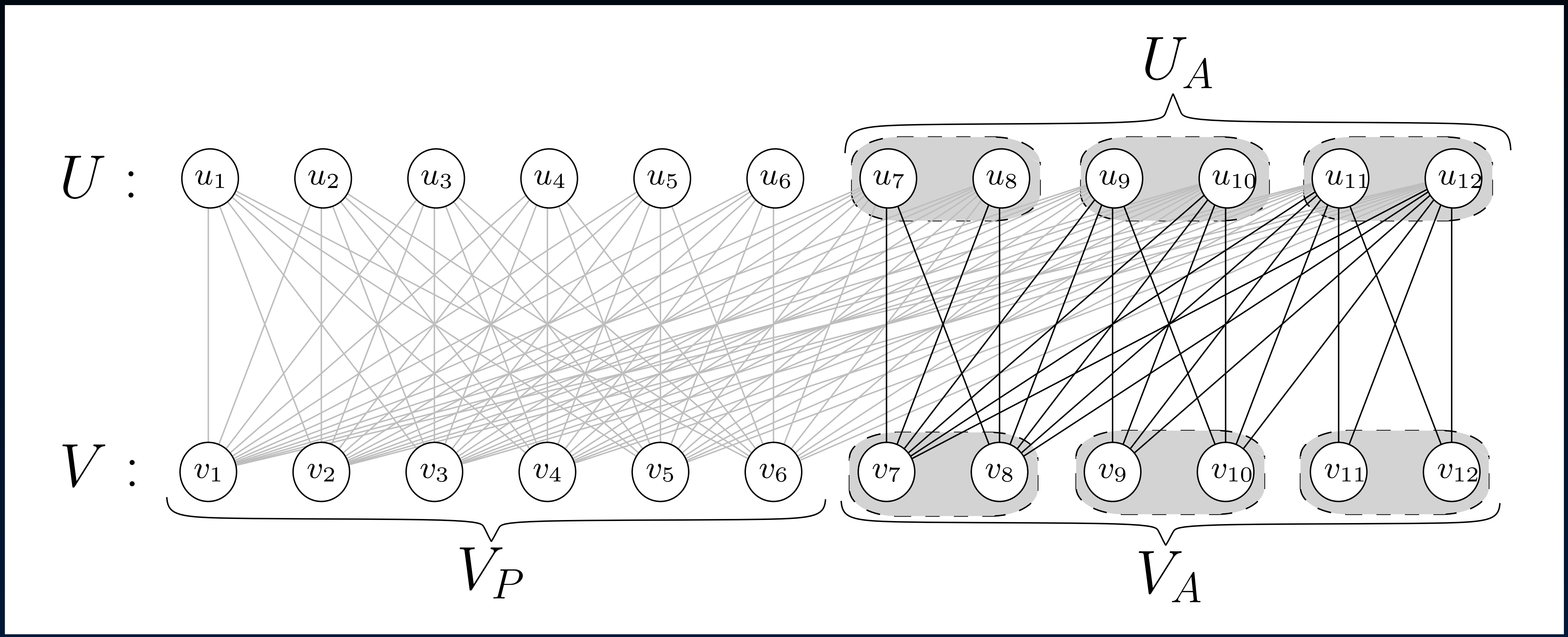
[1] Rajiv Gandhi, Samir Khuller, Srinivasan Parthasarathy, Aravind Srinivasan. Dependent rounding and its applications to approximation algorithms. JACM 53(3):324–360, 2006.

Marked nodes

- Adversary's goal:
  - Mark δn nodes in U whose complement has a matching in H
  - Minimize overlap with reserved nodes
- Best strategy:
  - Select $d_i = |T_i| - |S_i|$ nodes per component i
- $\mathbb{E}[|Reserved \cap Marked|] = \sum_i d_i \cdot \frac{d_i}{|T_i|} \geq \frac{(\delta n)^2}{n}$ (by Cauchy-Schwarz)
- $\Rightarrow 1 - \delta + \delta^2(1 - 1/e)$ competitive ratio



Hardness bound



- Predicted: complete graph; adversarial: block upper triangular
- Hardness of $1 - \delta e^{-\delta}$

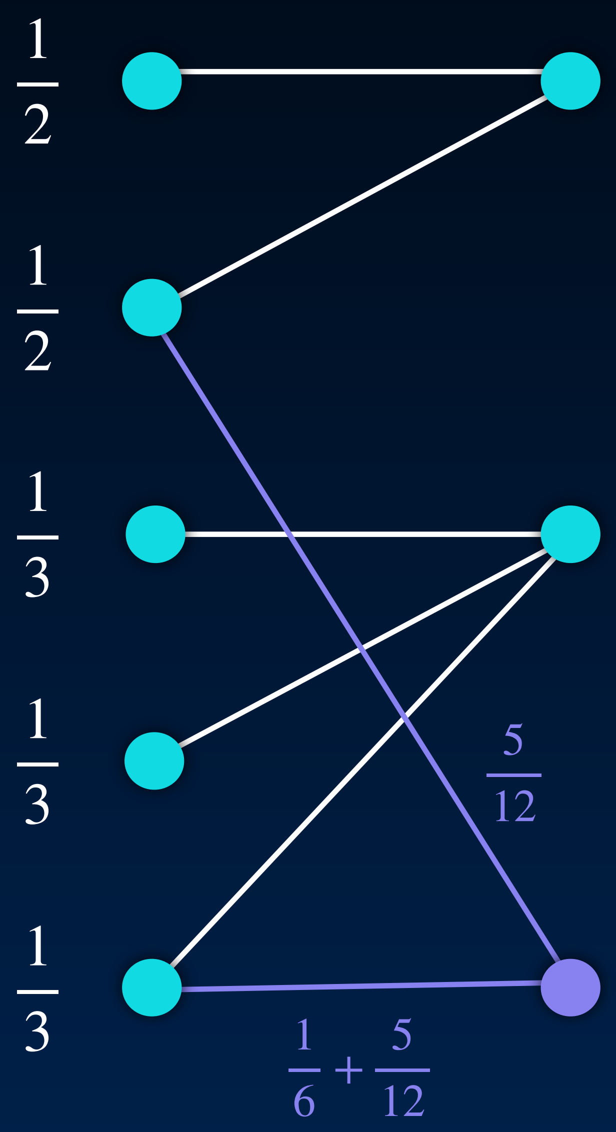
Fractional matching

- Online model
  - Nodes of V arrive one at a time, have to be fractionally matched to U
  - Water-level algorithm [1] gives optimal $1 - 1/e$ ratio
    - Match to the neighbor with lowest existing amount
- Semi-online fractional bipartite matching
  - We get tight bounds of $1 - \delta e^{-\delta}$
  - Primal-dual analysis extension of [2]

[1] Bala Kalyanasundaram and Kirk Pruhs.
An optimal deterministic algorithm for online b-matching. Theor. Comput. Sci., 233(1-2):319–325, 2000
[2] Nikhil R. Devanur, Kamal Jain, Robert D. Kleinberg.
Randomized primal-dual analysis of RANKING for online bipartite matching. SODA 2013

Algorithm for semi-online fractional matching

- For predicted nodes V_P:
  - Take fractional matching from skeleton decomposition of H
- For adversarial nodes V_A:
  - Use water-level algorithm



Primal-dual analysis

$$\max \sum_{e \in E} x_e$$

s.t.

$$\forall u \in U, \quad \sum_{v \in \delta(u)} x_{\{u,v\}} \leq 1$$

$$\forall v \in V, \quad \sum_{u \in \delta(v)} x_{\{u,v\}} \leq 1$$

$$\forall E \in E, \quad x_e \geq 0$$

$$\min \sum_{u \in U} \alpha_u + \sum_{v \in V} \beta_v$$

s.t.

$$\forall \{u,v\} \in E, \quad \alpha_u + \beta_v \geq 1$$

$$\forall u \in U, \quad \alpha_u \geq 0$$

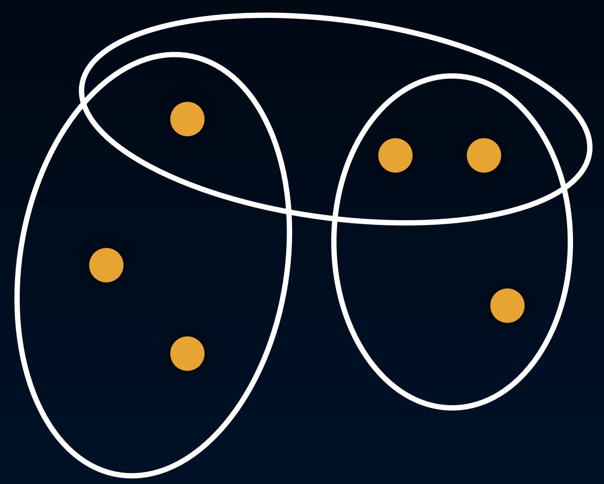
$$\forall v \in V, \quad \beta_v \geq 0$$

- For x found by our algorithm, set α_u and β_v such that
  - primal objective = dual objective
  - $\alpha_u + \beta_v \geq 1 - \delta e^{-\delta}$ for all edges

Summary

- Semi-online bipartite matching
  - Algorithm: $1 - \delta + \delta^2(1 - 1/e)$
  - Hardness: $1 - \delta e^{-\delta}$
  - Open problem: close the gap
- Fractional case
  - Algorithm and hardness: $1 - \delta e^{-\delta}$

Sets puzzle



- Ground set with n elements
- Collection of sets $\mathcal{S}$
  - Each $S \in \mathcal{S}$ contains d elements of $[n]$
- Player 1: pick $A \in \mathcal{S}$, maximize $|A \cap B|$
- Player 2: pick $B \in \mathcal{S}$, minimize $|A \cap B|$
- Show: there is a randomized strategy for player 1 to guarantee $\mathbb{E}[\,|A \cap B|\,] \geq d^2/n$