

Online Stochastic Matching Problem

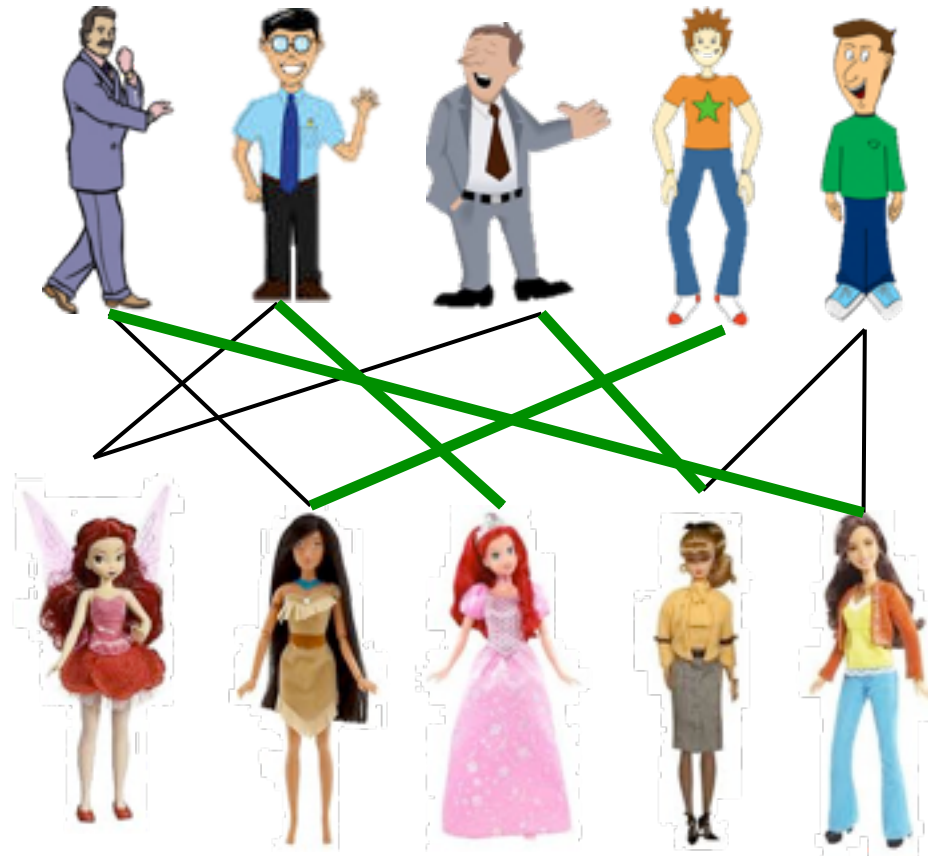
Shayan Oveis Gharan

Joint work with Vahideh Manshadi, Amin Saberi

Online Bipartite Matching Problem

Given a set of girls,
Boys arrive online,

Match each boy, irrevocably, maximizing the size of the matching.



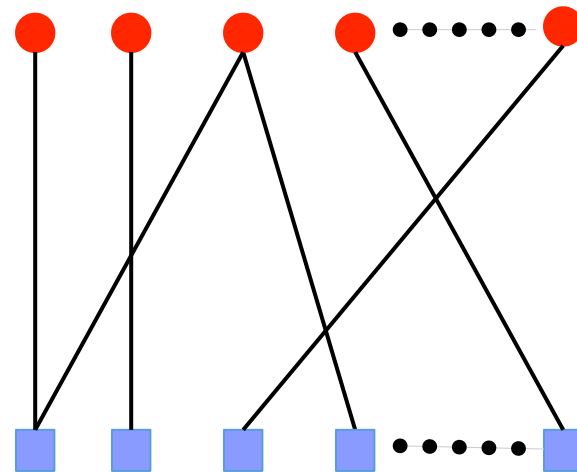
Online Bipartite Matching Problem (cont')

Given the set of bins,
Balls: adversarially/stochastically

Measure the performance against optimum (offline) solution

Balls arrive online

Bins are given



$$\text{Competitive Ratio} = \frac{E[ALG]}{E[OPT]}$$

Applications: Display Ads

Demand is offline from advertisers

Targeting a quantity ("1M ads a day")

Supply is online based on page views

The screenshot shows the Yahoo! homepage interface. At the top left is the 'YAHOO!' logo. To its right is a search bar with a yellow 'Search' button. Below the logo, the date 'Thursday, March 15, 2012' is displayed. On the right side, there are links for 'SIGN IN' (with 'New here? Sign Up') and 'MAIL' (with 'Check email').

On the left side, there is a 'YAHOO! SITES' menu with icons for Autos, Dating, Finance (Dow), Flickr, Games, Horoscopes, Jobs, Mail, Messenger, Movies, My Yahoo!, News, omg!, Real Estate, Screen, and Shine.

The main content area features a large video player with a 'Go to Video' button. Below the video is a news article titled 'Teen killed at party inspired by movie' with a sub-headline 'A free-for-all in a Houston mansion that was meant to emulate "Project X" turns deadly. It drew 500 to 1,000 guests >>'. To the right of the article is a 'TRENDING NOW' list:

- 01 Leah Ramini fired
- 02 Elvis granddaughter
- 03 Gallagher heart attack
- 04 Clint Eastwood
- 05 Tiger Woods
- 06 Ashley Judd
- 07 Brown recluse spider
- 08 Heart disease
- 09 Amanda Knox ex-b...
- 10 Foreclosures

Below the article are several small thumbnail images with captions: 'Gaga's gross confession', 'J.Lo's male look-alike', 'Death at wild teen party', 'Best shamrock shake recipe', and 'World's ugliest dog dies'. A pagination bar shows '1 - 5 of 60'.

At the bottom of the main content area is a 'March Madness 2012' section sponsored by 'PROGRESSIVE'. It displays a grid of team names and scores:

NMSU (13)	57	LOYM (15)	40	COLO (11)	36	CONN (5)	64
IND (4)	71	OHST (2)	57	UNLV (6)	25	IAST (8)	77

On the right side of the page, there is a large display advertisement for the '2012 Nissan Altima Sedan'. The ad features a silver car and lists incentives: '\$1,500 CASH BACK', '\$750 NMAC CASH', and '\$900 30th ANNIVERSARY PACKAGE SAVINGS', totaling '\$3,150 IN TOTAL SAVINGS'. A 'Shop Now' button is at the bottom left of the ad, and a 'NOW' logo is at the bottom right. A 'Visit NissanUSA.com - Ad Feedback' link is at the very bottom.

Models / Results

Models:

- **Adversarial**: Balls arrive adversarially.
- **Unknown Dist**: Balls are sampled i.i.d. from an unknown dist.
- **Known Dist**: Balls are sampled i.i.d. from a given dist.

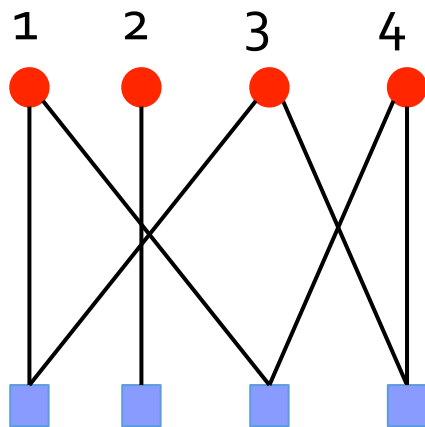
Model	Lower Bound	Upper Bound
Adversarial	$1-1/e$ (KVV'90-BM)	$1-1/e$ (KVV'90)
Known Dist. (integral rates)	0.677 (FMMM'09) 0.699 (BK'10), 0.705 (MOS'10) 0.729 (JX'11)	0.86 (MOS'10)
Unknown Dist.	0.655 (KMT'11) 0.696 (MY'11)	0.823 (MOS'10)

Online Stochastic Matching Problem (Known Dist.)

Given G with n ball types.

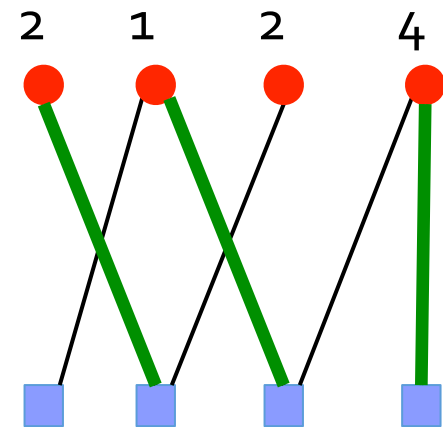
For $t=1, \dots, n$, select a ball **uniformly, with replacement** from the set of types.

Assign the ball to an empty bin, maximizing the expected number of assigned balls.



Expected graph G

Core Difficulty:
oversampling/
undersampling



Sample graph $G(\omega)$

Outline

- Rounding by Sampling
- Applying Rounding by Sampling to St. Matching Problem
 - Offline Statistics
- A More Adaptive Algorithm
- Open Problems/Future Works

Rounding By Sampling

Given a fractional vector f in an integral polytope.

Goal: Round f to an integer point

Write f as a convex combination of integer points

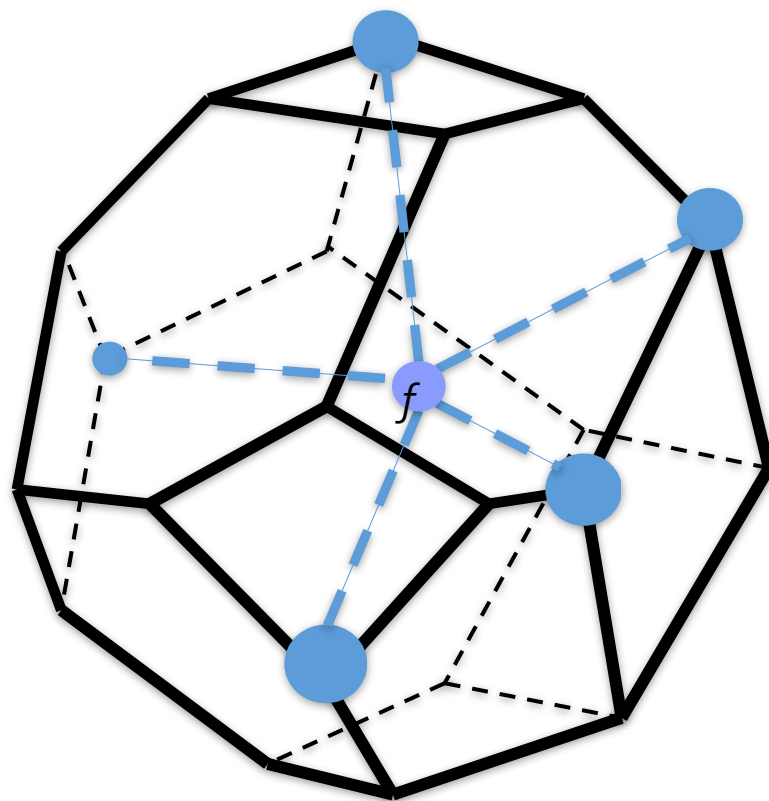
$$f = p_1 M_1 + p_2 M_2 + \dots + p_k M_k$$

Pick M_i with probability p_i

$$\Pr_{M \sim D} [e \in M] = f(e)$$

For any cost function $c(\cdot)$

$$E_{M \sim D} [c(M)] = c(f).$$



Additional Features of Particular Distributions

Distribution	Properties	Domain
Pipage Rounding	Negative Correlaion	Matroid Polytopes
Maximum Entropy	Strongest form of Negative Dep.	Spanning Trees
Randomized Swap Rounding	Negative Correlation, Lower tail bounds for Monotone Submodular f	Matroid Polytopes

Applications in Approximation Algorithms

- Maximizing a Submodular Function w.r.t. Matroid Constraint [Calenscu, Checkuri, Pal, Vondrak 07, Vondrak 08]
 - Pipage Rounding
- Asymmetric TSP [Asadpour, Goemans, Madry, O., Saberi'09]
 - Maximum Entropy dist.
- Symmetric TSP [O. Saberi, Singh'10]
 - Maximum Entropy dist.
- TSP Path [An, Kleinberg, Shmoys'11]
 - Any dist.

Outline

- Rounding by Sampling
- Applying Rounding by Sampling to St. Matching Problem
 - Offline Statistics
- A More Adaptive Algorithm
- Open Problems/Future Works

Offline Statistics

Let $f: E \rightarrow [0,1]$ be the expected optimum offline solution

$$f((x,y)) = \Pr_{G(\omega) \sim G} [x \text{ is matched to } y \text{ in the optimum}]$$

$$E[\text{OPT}] = \sum_{(x,y)} f(x,y)$$

f is a fractional matching, since

- Each ball type is sampled once in expectation
- Each bin is allocated at most once in each $G(\omega)$

Note: f can be estimated in polynomial time ...

Using Offline Statistics



Fractional
matching

Rounding by
Sampling

Integral
matchings

Rounding by Sampling 1 Matching

Offline:

Estimate $f(\cdot)$, Compute D

Sample a matching $M \sim D$.

Online:

Allocate first ball of each type to its match in M , and drop the rest

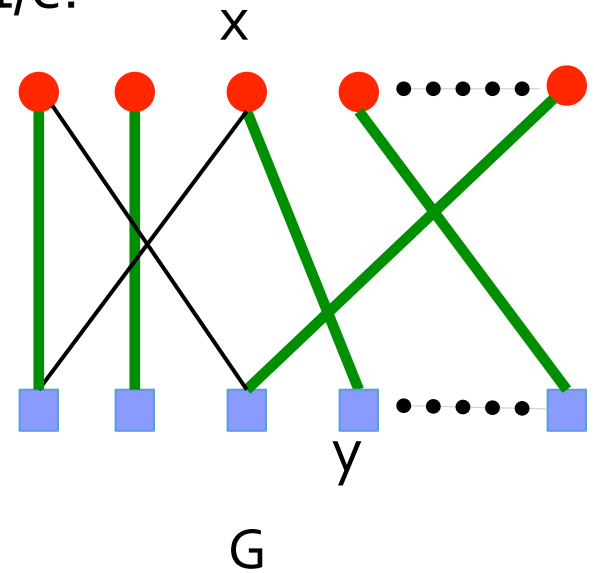
The algorithm has a competitive ratio of $1-1/e$:

$$E[\text{ALG}] \geq (1-1/e) |M|$$

For all, $(x,y) \in M$:

$$\begin{aligned} P[y \text{ is allocated}] &= P[x \text{ is sampled}] \\ &= 1 - (1-1/n)^n = 1-1/e \end{aligned}$$

$$E[\text{OPT}] = \sum_e f(e) = E[|M|]$$



Rounding by Sampling Two Matchings

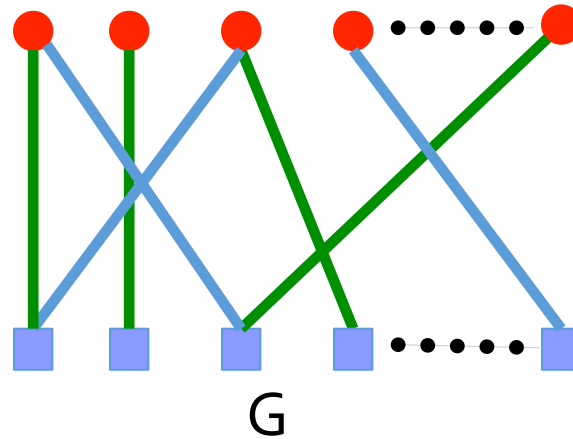
Offline:

Sample M_1 and M_2 independently from D

Online:

Allocate the first ball of each type according to M_1

Allocate the second ball of each type according to M_2



Lower Bounding ALG

Offline:

Sample M_1 and M_2 independently from D

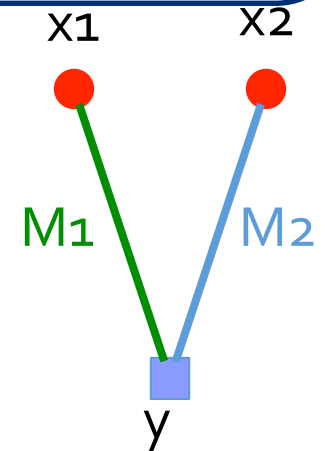
Online:

Allocate the first ball according to M_1

Allocate the second ball according to M_2

Algorithm, in expectation matches

$$(1-1/e)|M_1| + 1/e (1-2/e) |M_2 \setminus M_1|$$



$$\begin{aligned} P[y \text{ is allocated}] &= P[x_1 \text{ sampled}] + \\ &\quad + P[x_1 \text{ not sampled}] \cdot P[x_2 \text{ sampled twice} \mid x_1 \text{ not sampled}] \\ &= (1-1/e) + 1/e (1-2/e) \end{aligned}$$

Lower Bounding $E[|M_1|], E[|M_2|]$

Theorem [Manshadi, O, Saberi'10]: the algorithm has a competitive ratio of 0.668

Proof. $E[|M_1|] = \sum_e f(e) = E[OPT]$

$$E[|M_2 \setminus M_1|] = \sum_{(x,y)} \Pr[(x,y) \in M_2, (x,y) \notin M_1]$$

$$= \sum_{(x,y)} f(x,y) (1-f(x,y))$$

By independence

$$\geq E[OPT]/e$$

$$f(x,y) \leq \Pr[x \text{ sampled once}] = 1 - 1/e$$

$E[ALG]$

$E[OPT]$



Outline

- Rounding by Sampling
- Applying Rounding by Sampling to St. Matching Problem
 - Offline Statistics
- A More Adaptive Algorithm
- Open Problems/Future Works

A More Adaptive Algorithm

We can make the algorithm more adaptive by revising the decisions if the chosen bin is allocated.

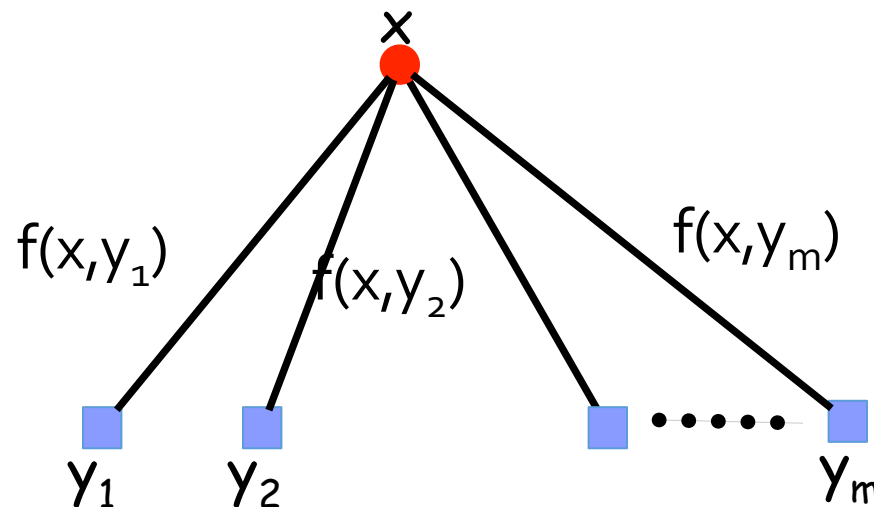
Offline: Estimate optimum offline solution $f(\cdot)$

Online: When a ball x arrives,

For $i=1 \rightarrow T$,

Sample $y \sim f(x, \cdot)$

If y is empty, assign x to y .



A More Adaptive Algorithm

Offline: Estimate optimum offline solution $f(\cdot)$, and dist. D

Online: When a ball x arrives,

For $i=1 \rightarrow T$,

 Sample $y \sim f(x, \cdot)$

 If y is empty, assign x to y .

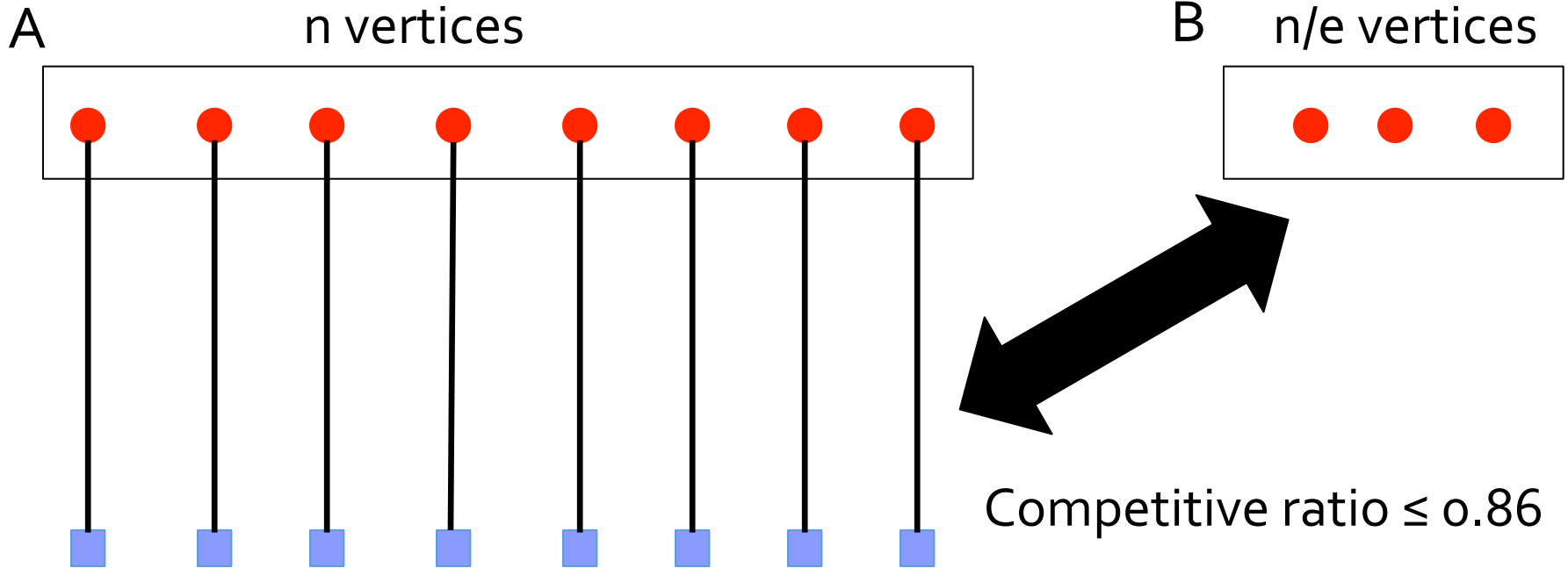
Theorem [MOS'10]: A variant of the above algorithm for $T=2$ has a 0.705 competitive ratio.

Theorem [JX'11]: A variant of the above algorithm for $T=3$ has a 0.729 competitive ratio.

Hard Example

OPT allocates all of the n bins

But, any online algorithm allocates at most $n(1-1/e^2)$.



Difficulty: n/e of the ball types in A will not be sampled at all
But, we do not know that until the end of the input.

Outline

- Rounding by Sampling
- Applying Rounding by Sampling to St. Matching Problem
 - Offline Statistics
- A More Adaptive Algorithm
- Open Problems/Future Works

BIG OPEN PROBLEM: Competing with Online Optimum

Since the arrival distribution is known, using Dynamic Programming,

Optimum Online algorithm is computable in time $O(n \cdot 2^n)$.

- How well can the optimum online be approximated?
- Does it admit a PTAS?

Future Works 1: Online Submodular Welfare Problem

Stochastic Matching problem as a special case of online stochastic SWP:

- $1-1/e$ is achievable for online stochastic SWP [Devanur, Jain, Sivan, Wilkens'11]
- There are submodular functions s.t. online SWP can not be approximated better than $1-1/e$ [Mirrokni, Shapira, Vondrak'o8]
- What about the functions that can be approximated efficiently?

[Haeupler, Mirrokni, Zadimoghaddam'11] obtained 0.667 for weighted matching problem

Future Works 2:k-Strongly Connected Subgraph problem

SCSP: Given a directed graph G , find the smallest k strongly connected subgraph of G .

A union of in-arborescence and out-arborescence is a 2-app.

[Laekhanukit, O., Singh'12]: Sampling the in-arborescence and out-arborescence independently based on LP solution gives a $1+1/k$.

- Obtaining better than 2 for the weighted version?

Conclusion and Open Problems

Model	Lower Bound	Upper Bound
Known Dist. With integer rates	0.729 (JX'11)	0.86 (MOS'10)
Known Dist.	0.708 (JX'11)	0.823 (MOS'10)
Known Dist. SWP	$1 - 1/e$ (DJSV'11)	$1 - 1/e$ (MSV'o8)