# Towards Optimal Degree Distributions for Left-perfect Matchings in Random Bipartite Graphs

Martin Dietzfelbinger*     Michael Rink

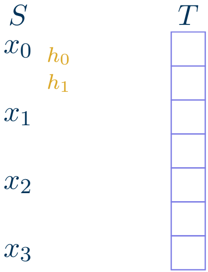
TECHNISCHE UNIVERSITÄT
ILMENAU

# Introduction

# Motivation

multiple choice hash table

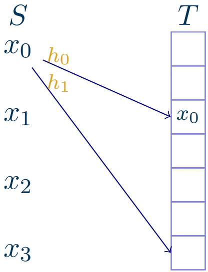
- set $S$ with $n$ keys, table $T$ with $m$ cells
- hash functions $h_0, h_1, \ldots, h_{d-1}$



# Motivation

multiple choice hash table

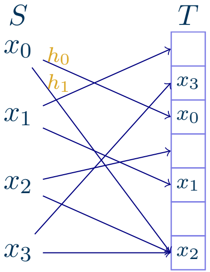
- set $S$ with $n$ keys, table $T$ with $m$ cells
- hash functions $h_0, h_1, \ldots, h_{d-1}$
- each key $x$ from $S$ is placed in one of $d \geq 2$ possible cells



# Motivation

multiple choice hash table

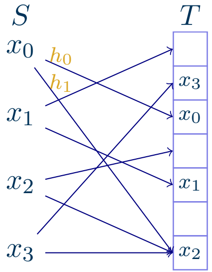
- set $S$ with $n$ keys, table $T$ with $m$ cells
- hash functions $h_0, h_1, \ldots, h_{d-1}$
- each key $x$ from $S$ is placed in one of $d \geq 2$ possible cells



# Motivation

multiple choice hash table

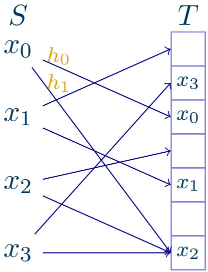
- set $S$ with $n$ keys, table $T$ with $m$ cells
- hash functions $h_0, h_1, \ldots, h_{d-1}$
- each key $x$ from $S$ is placed in one of $d \geq 2$ possible cells
- expected lookup time $(d+1)/2$
- here: static case — $S$ is known



# Motivation

multiple choice hash table

- set $S$ with $n$ keys, table $T$ with $m$ cells
- hash functions $h_0, h_1, \ldots, h_{d-1}$
- each key $x$ from $S$ is placed in one of $d \geq 2$ possible cells
- expected lookup time $(d+1)/2$
- here: static case — $S$ is known



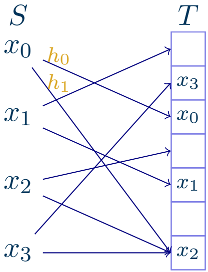

choice of $d$ is trade-off between

- space utilization (for $n = c \cdot m$ sharp thresholds are known)
- expected lookup time (critical if e.g. access to $T$ is slow)

# Motivation

multiple choice hash table

- set $S$ with $n$ keys, table $T$ with $m$ cells
- hash functions $h_0, h_1, \ldots, h_{d-1}$
- each key $x$ from $S$ is placed in one of $d \geq 2$ possible cells
- expected lookup time $(d+1)/2$
- here: static case — $S$ is known




choice of $d$ is trade-off between

- space utilization (for $n = c \cdot m$ sharp thresholds are known)
- expected lookup time (critical if e.g. access to $T$ is slow)

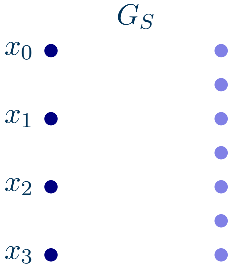
**Question**

Can we do better if the keys are not restricted to use the same number of possible table cells?

# Graph Model

bipartite graph $G_S$

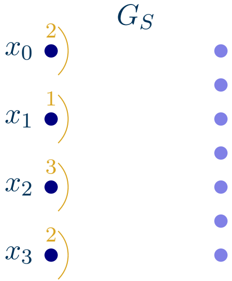
- $n$ left nodes, labeled with elements from $S$, $m$ right nodes
- for each $x$ from $S$ the degree is a random variable $D_x$



# Graph Model

bipartite graph $G_S$

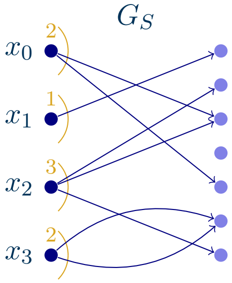
- $n$ left nodes, labeled with elements from $S$, $m$ right nodes
- for each $x$ from $S$ the degree is a random variable $D_x$
  - ★ randomly choose its degree $d$



# Graph Model

bipartite graph $G_S$

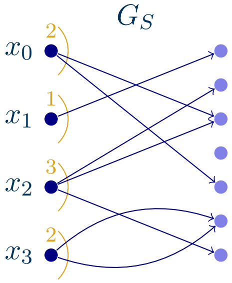
- $n$ left nodes, labeled with elements from $S$, $m$ right nodes
- for each $x$ from $S$ the degree is a random variable $D_x$
  - randomly choose its degree $d$
  - choose $d$ random right neighbors



# Graph Model

bipartite graph $G_S$

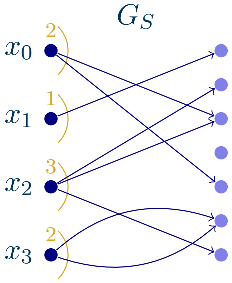
- $n$ left nodes, labeled with elements from $S$, $m$ right nodes
- for each $x$ from $S$ the degree is a random variable $D_x$
  - ★ randomly choose its degree $d$
  - ★ choose $d$ random right neighbors
  - ★ all choices are independent



# Graph Model

bipartite graph $G_S$

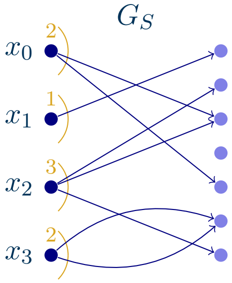
- $n$ left nodes, labeled with elements from $S$, $m$ right nodes
- for each $x$ from $S$ the degree is a random variable $D_x$
  - ★ randomly choose its degree $d$
  - ★ choose $d$ random right neighbors
  - ★ all choices are independent
- each $D_x$ follows its own probability mass function $\rho_x$



# Graph Model

bipartite graph $G_S$

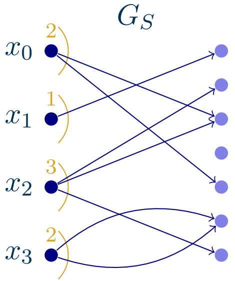
- $n$ left nodes, labeled with elements from $S$, $m$ right nodes
- for each $x$ from $S$ the degree is a random variable $D_x$
  - ⋆ randomly choose its degree $d$
  - ⋆ choose $d$ random right neighbors
  - ⋆ all choices are independent
- each $D_x$ follows its own probability mass function $\rho_x$
- $\Delta_x$ is the expectation of $D_x$
- $\bar{\Delta}$ is the average expected value



# Graph Model

bipartite graph $G_S$

- $n$ left nodes, labeled with elements from $S$, $m$ right nodes
- for each $x$ from $S$ the degree is a random variable $D_x$
  - ⋆ randomly choose its degree $d$
  - ⋆ choose $d$ random right neighbors
  - ⋆ all choices are independent
- each $D_x$ follows its own probability mass function $\rho_x$
- $\Delta_x$ is the expectation of $D_x$
- $\bar{\Delta}$ is the average expected value




**Question**

Given parameters $(n, m, \bar{\Delta})$, which sequence of probability mass functions $(\rho_x)_{x \in S}$ maximizes the probability that $G_S$ has a left-perfect matching?

# Main Result

**Theorem**

Let $n \leq m$, as well as $n, \bar{\Delta} \geq 2$, and let $(\rho_x)_{x \in S}$ be an optimal sequence for parameters $(n, m, \bar{\Delta})$. Then the following holds for all $x \in S$.

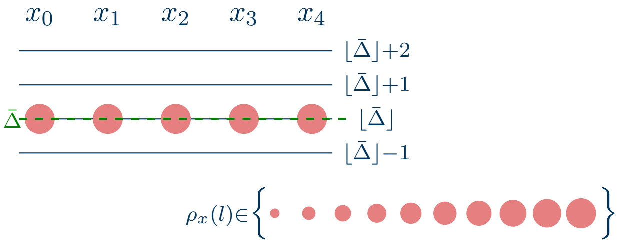
# Main Result

**Theorem**

Let $n \leq m$, as well as $n, \bar{\Delta} \geq 2$, and let $(\rho_x)_{x \in S}$ be an optimal sequence for parameters $(n, m, \bar{\Delta})$. Then the following holds for all $x \in S$.

- If $\bar{\Delta}$ is an integer, then $\rho_x(\bar{\Delta}) = 1$.

$\bar{\Delta}$ is an integer

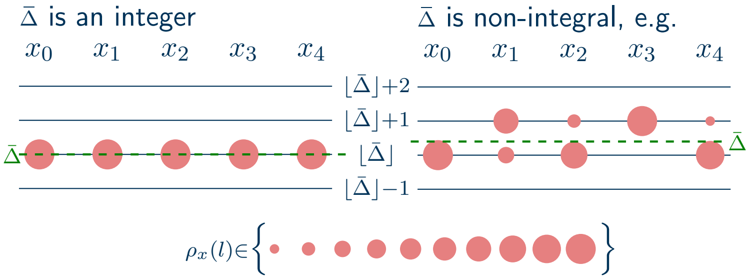


# Main Result

**Theorem**

Let $n \leq m$, as well as $n, \bar{\Delta} \geq 2$, and let $(\rho_x)_{x \in S}$ be an optimal sequence for parameters $(n, m, \bar{\Delta})$. Then the following holds for all $x \in S$.

- If $\bar{\Delta}$ is an integer, then $\rho_x(\bar{\Delta}) = 1$.
- If $\bar{\Delta}$ is non-integral, then $\rho_x(\lfloor\bar{\Delta}\rfloor) \in [0, 1]$ and $\rho_x(\lceil\bar{\Delta}\rceil) = 1 - \rho_x(\lfloor\bar{\Delta}\rfloor)$.



# Related Work

generalized Cuckoo Hashing (similar graphs, pairwise distinct neighbors)

- thresholds for regular case — 1 fixed degree
  - ⋆ [Frieze and Melsted, 2009]
  - ⋆ [Fountoulakis and Panagiotou, 2010]
  - ⋆ [D, Goerdt, Mitzenmacher, Montanari, Pagh, R, 2010]

# Related Work

generalized Cuckoo Hashing (similar graphs, pairwise distinct neighbors)

- thresholds for regular case — 1 fixed degree
  - ⋆ [Frieze and Melsted, 2009]
  - ⋆ [Fountoulakis and Panagiotou, 2010]
  - ⋆ [D, Goerdt, Mitzenmacher, Montanari, Pagh, R, 2010]
- thresholds for irregular case — 2 different degrees
  + variant of Lemma 1 (next slide)
  - ⋆ [D, Goerdt, Mitzenmacher, Montanari, Pagh, R, 2010]

# Related Work

generalized Cuckoo Hashing (similar graphs, pairwise distinct neighbors)

- thresholds for regular case — 1 fixed degree
  - ⋆ [Frieze and Melsted, 2009]
  - ⋆ [Fountoulakis and Panagiotou, 2010]
  - ⋆ [D, Goerdt, Mitzenmacher, Montanari, Pagh, R, 2010]
- thresholds for irregular case — 2 different degrees
  + variant of Lemma 1 (next slide)
  - ⋆ [D, Goerdt, Mitzenmacher, Montanari, Pagh, R, 2010]

erasure correcting codes (different graphs)

# Related Work

generalized Cuckoo Hashing (similar graphs, pairwise distinct neighbors)

- thresholds for regular case — 1 fixed degree
  - [Frieze and Melsted, 2009]
  - [Fountoulakis and Panagiotou, 2010]
  - [D, Goerdt, Mitzenmacher, Montanari, Pagh, R, 2010]
- thresholds for irregular case — 2 different degrees + variant of Lemma 1 (next slide)
  - [D, Goerdt, Mitzenmacher, Montanari, Pagh, R, 2010]

erasure correcting codes (different graphs)

- uneven distribution increases probability for existence of a matching that can be calculated in linear time (peeling)
  - Tornado Codes [Luby, Mitzenmacher, Shokrollahi, Spielman, 2001]
  - LT Codes [Luby, 2002]
  - Online Codes [Maymounkov, 2002]
  - Raptor Codes [Shokrollahi, 2006]

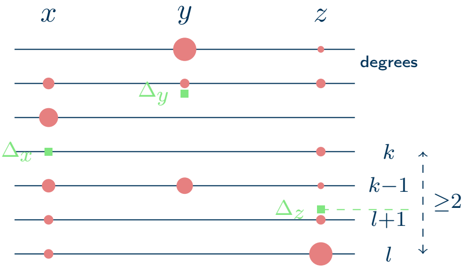
  ...

# Proof of Main Theorem

# Degrees Must be Next to the Mean

**Lemma**

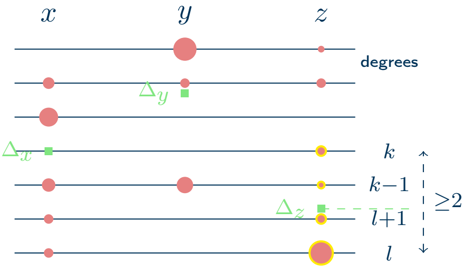
Let $(\rho_x)_{x\in S}$ be given. Let $z \in S$ be arbitrary but fixed. If in $\rho_z$ two degrees with distance at least $2$ have nonzero probability then $(\rho_x)_{x\in S}$ is not optimal.



# Degrees Must be Next to the Mean

**Lemma**

Let $(\rho_x)_{x \in S}$ be given. Let $z \in S$ be arbitrary but fixed. If in $\rho_z$ two degrees with distance at least $2$ have nonzero probability then $(\rho_x)_{x \in S}$ is not optimal.



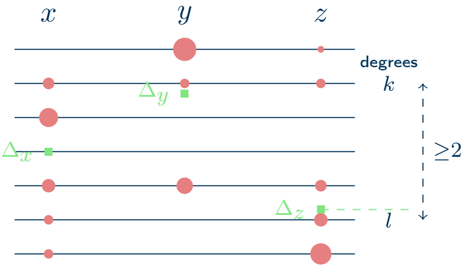

For some $\varepsilon > 0$

- increase the probability of degrees $l+1$ and $k-1$ by $\varepsilon$
- decrease the probability of degrees $l$ and $k$ by $\varepsilon$

# Degrees Must be Next to the Mean

**Lemma**

Let $(\rho_x)_{x\in S}$ be given. Let $z \in S$ be arbitrary but fixed. If in $\rho_z$ two degrees with distance at least $2$ have nonzero probability then $(\rho_x)_{x\in S}$ is not optimal.




For some $\varepsilon > 0$

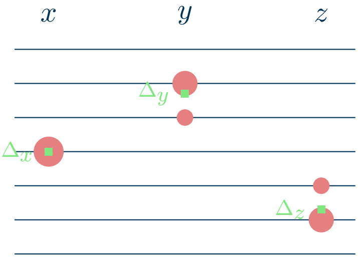
- increase the probability of degrees $l+1$ and $k-1$ by $\varepsilon$
- decrease the probability of degrees $l$ and $k$ by $\varepsilon$

This decreases the failure probability but does not change $\Delta_z$ and $\bar{\Delta}$.

# Degrees Must be Next to the Mean

**Lemma**

Let $(\rho_x)_{x \in S}$ be given. Let $z \in S$ be arbitrary but fixed. If in $\rho_z$ two degrees with distance at least $2$ have nonzero probability then $(\rho_x)_{x \in S}$ is not optimal.



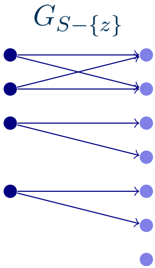

For some $\varepsilon > 0$

- increase the probability of degrees $l + 1$ and $k - 1$ by $\varepsilon$
- decrease the probability of degrees $l$ and $k$ by $\varepsilon$

This decreases the failure probability but does not change $\Delta_z$ and $\bar{\Delta}$.

# Proof Idea

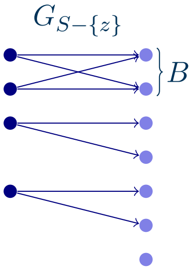
Fix graph without node $z$. $G_{S-\{z\}}$ may admit many matchings.
Subdivide right node set into



# Proof Idea

Fix graph without node $z$. $G_{S-\{z\}}$ may admit many matchings. Subdivide right node set into

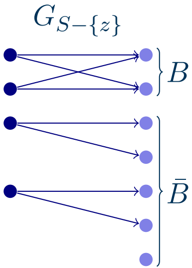
- $B$: blocked nodes, nodes that are matched in all matchings



# Proof Idea

Fix graph without node $z$. $G_{S-\{z\}}$ may admit many matchings.
Subdivide right node set into

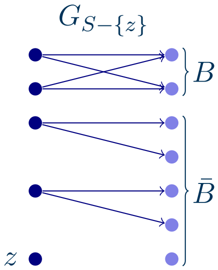
- $B$ : blocked nodes, nodes that are matched in all matchings
- $\bar{B}$ : non-blocked nodes



# Proof Idea

Fix graph without node $z$. $G_{S-\{z\}}$ may admit many matchings. Subdivide right node set into

- $B$: blocked nodes, nodes that are matched in all matchings
- $\bar{B}$: non-blocked nodes

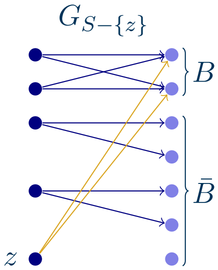



**$G_S$ has no matching iff**

# Proof Idea

Fix graph without node $z$. $G_{S-\{z\}}$ may admit many matchings. Subdivide right node set into

- $B$: blocked nodes, nodes that are matched in all matchings
- $\bar{B}$: non-blocked nodes



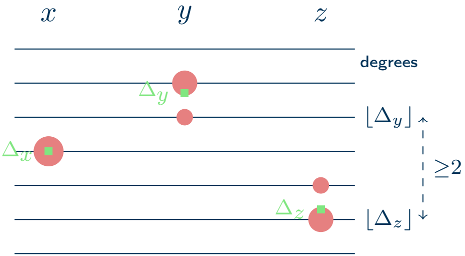

**$G_S$ has no matching iff**

$$N_z := \{u \mid (z,u) \text{ is edge in } G_S\} \subseteq B$$

# Average Degrees of Different Nodes are Close

**Lemma**

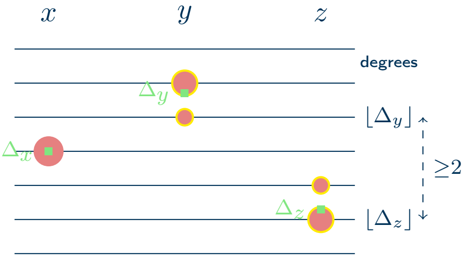
Let $(\rho_x)_{x \in S}$ be given, where for each $x \in S$ only degrees from $\{\lfloor \Delta_x \rfloor, \lceil \Delta_x \rceil\}$ have nonzero probability. Let $y, z \in S$ be arbitrary but fixed. If $\lfloor \Delta_y \rfloor$ and $\lfloor \Delta_z \rfloor$ have distance at least $2$, or $\lceil \Delta_y \rceil$ and $\lceil \Delta_z \rceil$ have distance at least $2$, then $(\rho_x)_{x \in S}$ is not optimal.



# Average Degrees of Different Nodes are Close

**Lemma**

Let $(\rho_x)_{x\in S}$ be given, where for each $x \in S$ only degrees from $\{\lfloor\Delta_x\rfloor, \lceil\Delta_x\rceil\}$ have nonzero probability. Let $y, z \in S$ be arbitrary but fixed. If $\lfloor\Delta_y\rfloor$ and $\lfloor\Delta_z\rfloor$ have distance at least $2$, or $\lceil\Delta_y\rceil$ and $\lceil\Delta_z\rceil$ have distance at least $2$, then $(\rho_x)_{x\in S}$ is not optimal.



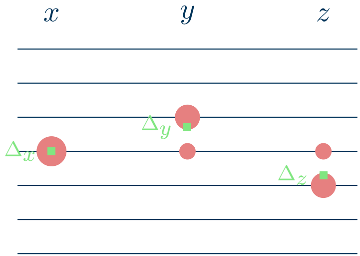

Decrease the failure probability by

- increasing $\Delta_z$ by $1$
- decreasing $\Delta_y$ by $1$

# Average Degrees of Different Nodes are Close

**Lemma**

Let $(\rho_x)_{x \in S}$ be given, where for each $x \in S$ only degrees from $\{\lfloor \Delta_x \rfloor, \lceil \Delta_x \rceil\}$ have nonzero probability. Let $y, z \in S$ be arbitrary but fixed. If $\lfloor \Delta_y \rfloor$ and $\lfloor \Delta_z \rfloor$ have distance at least $2$, or $\lceil \Delta_y \rceil$ and $\lceil \Delta_z \rceil$ have distance at least $2$, then $(\rho_x)_{x \in S}$ is not optimal.



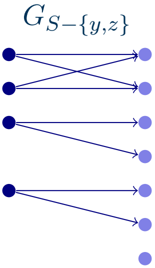

Decrease the failure probability by

- increasing $\Delta_z$ by $1$
- decreasing $\Delta_y$ by $1$

This does not change $\bar{\Delta}$.

# Proof Idea (1)

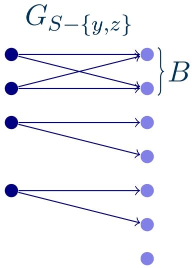
Fix graph without nodes $y$ and $z$. Subdivide right node set into



# Proof Idea (1)

Fix graph without nodes $y$ and $z$. Subdivide right node set into

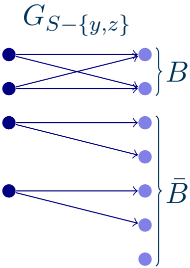
- $B$: blocked nodes, nodes that are matched in all matchings



# Proof Idea (1)

Fix graph without nodes $y$ and $z$. Subdivide right node set into

- $B$: blocked nodes, nodes that are matched in all matchings
- $\bar{B}$: non-blocked nodes



# Proof Idea (1)

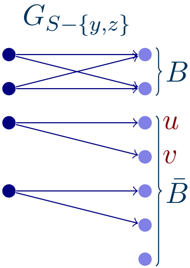
Fix graph without nodes $y$ and $z$. Subdivide right node set into

- $B$: blocked nodes, nodes that are matched in all matchings
- $\bar{B}$: non-blocked nodes

Define relation $\sim$ for nodes from $\bar{B}$:

**$u \sim v$ iff**

there is no matching of $G_{S-\{y,z\}}$ that leaves $u$ and $v$ simultaneously unmatched



# Proof Idea (1)

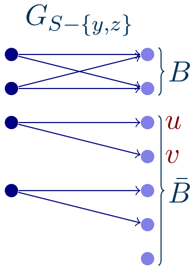
Fix graph without nodes $y$ and $z$. Subdivide right node set into

- $B$: blocked nodes, nodes that are matched in all matchings
- $\bar{B}$: non-blocked nodes

Define relation $\sim$ for nodes from $\bar{B}$:

**$u \sim v$ iff**

there is no matching of $G_{S-\{y,z\}}$ that leaves $u$ and $v$ simultaneously unmatched




Prove that $\sim$ is an equivalence relation.

# Proof Idea (1)

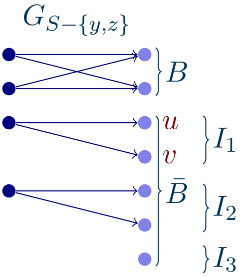
Fix graph without nodes $y$ and $z$. Subdivide right node set into

- $B$: blocked nodes, nodes that are matched in all matchings
- $\bar{B}$: non-blocked nodes

Define relation $\sim$ for nodes from $\bar{B}$:

**$u \sim v$ iff**

there is no matching of $G_{S-\{y,z\}}$ that leaves $u$ and $v$ simultaneously unmatched




Prove that $\sim$ is an equivalence relation.

Define sets $I_j$ as equivalence classes of $\sim$.

# Proof Idea (1)

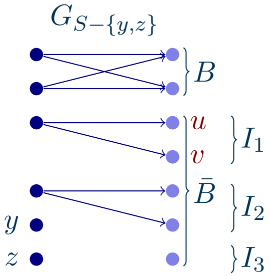
Fix graph without nodes $y$ and $z$. Subdivide right node set into

- $B$: blocked nodes, nodes that are matched in all matchings
- $\bar{B}$: non-blocked nodes

Define relation $\sim$ for nodes from $\bar{B}$:

**$u \sim v$ iff**

there is no matching of $G_{S-\{y,z\}}$ that leaves $u$ and $v$ simultaneously unmatched




Prove that $\sim$ is an equivalence relation.

Define sets $I_j$ as equivalence classes of $\sim$.

**$G_S$ has no matching iff**

# Proof Idea (1)

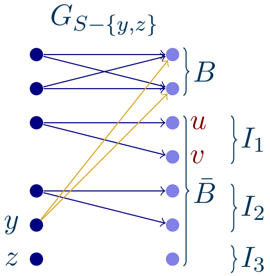
Fix graph without nodes $y$ and $z$. Subdivide right node set into

- $B$: blocked nodes, nodes that are matched in all matchings
- $\bar{B}$: non-blocked nodes

Define relation $\sim$ for nodes from $\bar{B}$:

**$u \sim v$ iff**

there is no matching of $G_{S-\{y,z\}}$ that leaves $u$ and $v$ simultaneously unmatched




Prove that $\sim$ is an equivalence relation.

Define sets $I_j$ as equivalence classes of $\sim$.

**$G_S$ has no matching iff**

$(N_y \subseteq B)$

# Proof Idea (1)

Fix graph without nodes $y$ and $z$. Subdivide right node set into

- $B$: blocked nodes, nodes that are matched in all matchings
- $\bar{B}$: non-blocked nodes

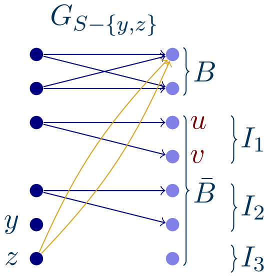
Define relation $\sim$ for nodes from $\bar{B}$:

**$u \sim v$ iff**

there is no matching of $G_{S-\{y,z\}}$ that leaves $u$ and $v$ simultaneously unmatched

Prove that $\sim$ is an equivalence relation.

Define sets $I_j$ as equivalence classes of $\sim$.




**$G_S$ has no matching iff**

$$(N_y \subseteq B) \vee (N_z \subseteq B)$$

# Proof Idea (1)

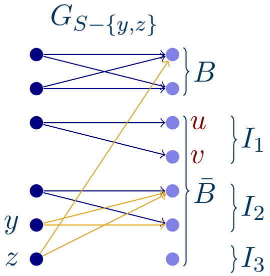
Fix graph without nodes $y$ and $z$. Subdivide right node set into

- $B$: blocked nodes, nodes that are matched in all matchings
- $\bar{B}$: non-blocked nodes

Define relation $\sim$ for nodes from $\bar{B}$:

**$u \sim v$ iff**

there is no matching of $G_{S-\{y,z\}}$ that leaves $u$ and $v$ simultaneously unmatched




Prove that $\sim$ is an equivalence relation.

Define sets $I_j$ as equivalence classes of $\sim$.

**$G_S$ has no matching iff**

$$(N_y \subseteq B) \vee (N_z \subseteq B) \vee \bigvee_j \Big( (N_y \cup N_z) \subseteq (B \cup I_j) \Big)$$

# Proof Idea (2)

Assume: $b$ blocked nodes, $r$ equivalence classes $I_j$ of size $i_j$

# Proof Idea (2)

Assume: $b$ blocked nodes, $r$ equivalence classes $I_j$ of size $i_j$

Failure probability for $D_y = d_y$ and $D_z = d_z$:

# Proof Idea (2)

Assume: $b$ blocked nodes, $r$ equivalence classes $I_j$ of size $i_j$

Failure probability for $D_y = d_y$ and $D_z = d_z$:

$$\text{fail}(d_y, d_z, b, r, i_1, \ldots, i_r) = \underbrace{\left(\frac{b}{m}\right)^{d_y}}_{N_y \subseteq B} + \underbrace{\left(\frac{b}{m}\right)^{d_z}}_{N_z \subseteq B} - \underbrace{\left(\frac{b}{m}\right)^{d_y} \cdot \left(\frac{b}{m}\right)^{d_z}}_{(N_y \cup N_z) \subseteq B}$$

$$+$$

# Proof Idea (2)

Assume: $b$ blocked nodes, $r$ equivalence classes $I_j$ of size $i_j$

Failure probability for $D_y = d_y$ and $D_z = d_z$:

$$\mathrm{fail}(d_y, d_z, b, r, i_1, \ldots, i_r) = \underbrace{\left(\frac{b}{m}\right)^{d_y}}_{N_y \subseteq B} + \underbrace{\left(\frac{b}{m}\right)^{d_z}}_{N_z \subseteq B} - \underbrace{\left(\frac{b}{m}\right)^{d_y} \cdot \left(\frac{b}{m}\right)^{d_z}}_{(N_y \cup N_z) \subseteq B}$$

$$+ \sum_{j=1}^{r} \left[ \underbrace{\left(\frac{i_j + b}{m}\right)^{d_y}}_{N_y \subseteq (B \cup I_j)} - \left(\frac{b}{m}\right)^{d_y} \right] \cdot \left[ \underbrace{\left(\frac{i_j + b}{m}\right)^{d_z}}_{N_z \subseteq (B \cup I_j)} - \left(\frac{b}{m}\right)^{d_z} \right]$$

# Proof Idea (2)

Assume: $b$ blocked nodes, $r$ equivalence classes $I_j$ of size $i_j$

Failure probability for $D_y = d_y$ and $D_z = d_z$:

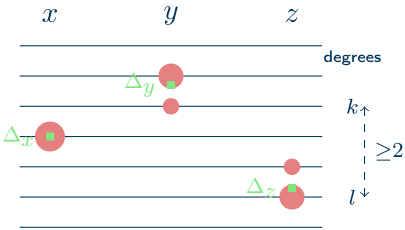
$$\text{fail}(d_y, d_z, b, r, i_1, \dots, i_r) = \underbrace{\left(\frac{b}{m}\right)^{d_y}}_{N_y \subseteq B} + \underbrace{\left(\frac{b}{m}\right)^{d_z}}_{N_z \subseteq B} - \underbrace{\left(\frac{b}{m}\right)^{d_y} \cdot \left(\frac{b}{m}\right)^{d_z}}_{(N_y \cup N_z) \subseteq B}$$

$$+ \sum_{j=1}^{r} \left[ \underbrace{\left(\frac{i_j + b}{m}\right)^{d_y}}_{N_y \subseteq (B \cup I_j)} - \left(\frac{b}{m}\right)^{d_y} \right] \cdot \left[ \underbrace{\left(\frac{i_j + b}{m}\right)^{d_z}}_{N_z \subseteq (B \cup I_j)} - \left(\frac{b}{m}\right)^{d_z} \right]$$

The failure probability decreases if

$$\text{fail}(k, l) > \text{fail}(k-1, l+1) \ ,$$

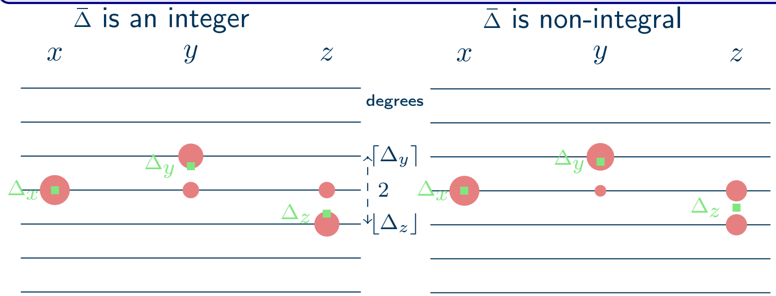
for each vector $(b, r, i_1, \dots, i_r)$.



# Optimal Distributions Use At Most Two Neighboring Degrees

**Lemma**

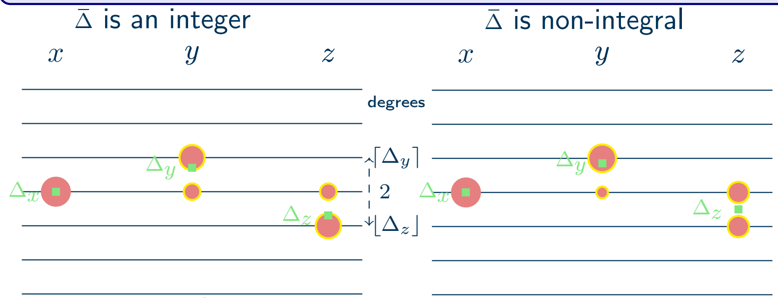
Let $(\rho_x)_{x\in S}$ be given as in the last lemma. Let $y, z \in S$ be arbitrary but fixed and assume that $\Delta_y$ and $\Delta_z$ are non-integral. If $\lceil \Delta_y \rceil$ and $\lfloor \Delta_z \rfloor$ have distance $2$ then $(\rho_x)_{x\in S}$ is not optimal.



# Optimal Distributions Use At Most Two Neighboring Degrees

**Lemma**

Let $(\rho_x)_{x \in S}$ be given as in the last lemma. Let $y, z \in S$ be arbitrary but fixed and assume that $\Delta_y$ and $\Delta_z$ are non-integral. If $\lceil \Delta_y \rceil$ and $\lfloor \Delta_z \rfloor$ have distance $2$ then $(\rho_x)_{x \in S}$ is not optimal.




For some perturbation $\varepsilon \neq 0$

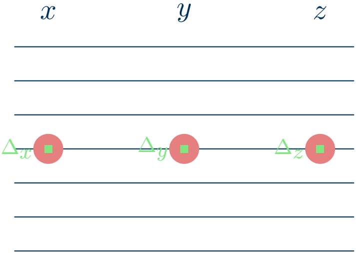
- increase the probability of degree $\lfloor \Delta_y \rfloor$ and $\lceil \Delta_z \rceil$ by $\varepsilon$
- decrease the probability of degree $\lceil \Delta_y \rceil$ and $\lfloor \Delta_z \rfloor$ by $\varepsilon$

# Optimal Distributions Use At Most Two Neighboring Degrees

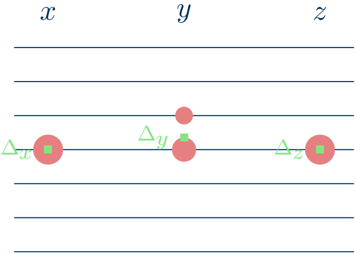
**Lemma**

Let $(\rho_x)_{x\in S}$ be given as in the last lemma. Let $y, z \in S$ be arbitrary but fixed and assume that $\Delta_y$ and $\Delta_z$ are non-integral. If $\lceil \Delta_y \rceil$ and $\lfloor \Delta_z \rfloor$ have distance $2$ then $(\rho_x)_{x\in S}$ is not optimal.

$\bar{\Delta}$ is an integer




$\bar{\Delta}$ is non-integral



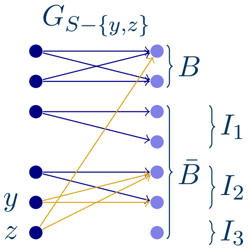

For some perturbation $\varepsilon \neq 0$

- increase the probability of degree $\lfloor \Delta_y \rfloor$ and $\lceil \Delta_z \rceil$ by $\varepsilon$
- decrease the probability of degree $\lceil \Delta_y \rceil$ and $\lfloor \Delta_z \rfloor$ by $\varepsilon$

This decreases the failure probability but does not change $\bar{\Delta}$.

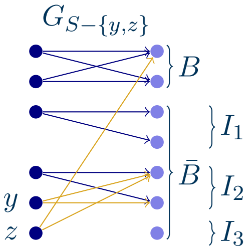
# Proof Idea

Same approach as before — $b$ blocked nodes, $r$ equivalence classes of size $i_j$



# Proof Idea

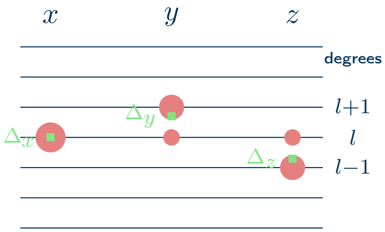
Same approach as before — $b$ blocked nodes, $r$ equivalence classes of size $i_j$




The failure probability decreases if

$$\mathrm{fail}(l,l) - \mathrm{fail}(l+1,l) < \mathrm{fail}(l-1,l) - \mathrm{fail}(l,l) \ ,$$

for each vector $(b, r, i_1, \ldots, i_r)$.



# Conjecture

# Non-integral Case

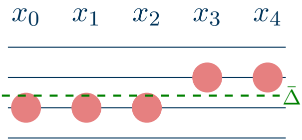
optimal distribution could be

sufficient condition

# Non-integral Case

optimal distribution could be

- fixed degree




sufficient condition

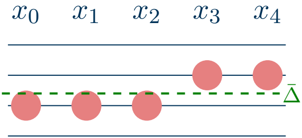
$$\mathrm{fail}(l,l) - \mathrm{fail}(l+1,l)$$
$$> \mathrm{fail}(l,l+1) - \mathrm{fail}(l+1,l+1)$$

(convex)

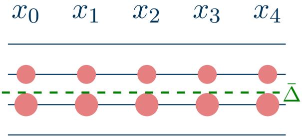
# Non-integral Case

optimal distribution could be

- fixed degree




- Binomial distribution




sufficient condition

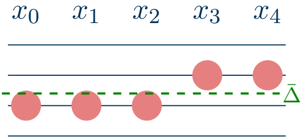
$$\mathrm{fail}(l,l) - \mathrm{fail}(l+1,l)$$
$$> \mathrm{fail}(l,l+1) - \mathrm{fail}(l+1,l+1)$$
(convex)

$$\mathrm{fail}(l,l) - \mathrm{fail}(l+1,l)$$
$$< \mathrm{fail}(l,l+1) - \mathrm{fail}(l+1,l+1)$$
(concave)

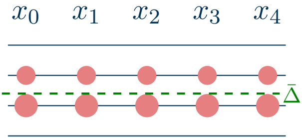
# Non-integral Case

optimal distribution could be

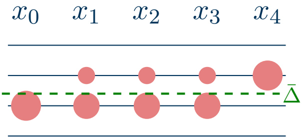
- fixed degree




- Binomial distribution




- other distribution, e.g.,




sufficient condition

$$\begin{aligned} & \text{fail}(l,l) - \text{fail}(l+1,l) \\ > \; & \text{fail}(l,l+1) - \text{fail}(l+1,l+1) \end{aligned}$$

(convex)

$$\begin{aligned} & \text{fail}(l,l) - \text{fail}(l+1,l) \\ < \; & \text{fail}(l,l+1) - \text{fail}(l+1,l+1) \end{aligned}$$

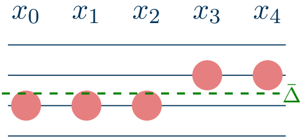
(concave)

?

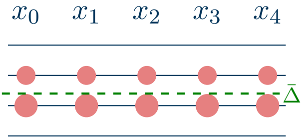
# Non-integral Case

optimal distribution could be

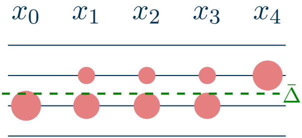
- fixed degree




- Binomial distribution




- other distribution, e.g.,




sufficient condition

$$\begin{aligned} & \mathrm{fail}(l,l) - \mathrm{fail}(l+1,l) \\ > \; & \mathrm{fail}(l,l+1) - \mathrm{fail}(l+1,l+1) \end{aligned}$$

(convex)

$$\begin{aligned} & \mathrm{fail}(l,l) - \mathrm{fail}(l+1,l) \\ < \; & \mathrm{fail}(l,l+1) - \mathrm{fail}(l+1,l+1) \end{aligned}$$

(concave)

?

Not easy to decide. Consider special case $n = \Theta(m)$.

# Existence of Thresholds

**Proposition**

Let $n = c \cdot m$, for constant $c > 0$, and let $(\rho_x)_{x \in S}$ be a near optimal sequence with average expected degree $\bar{\Delta} > 2$. Then for sufficiently large $n$ there is a threshold $c^*(\bar{\Delta})$ such that the random graph $G_S$ has the following property.

# Existence of Thresholds

**Proposition**

Let $n = c \cdot m$, for constant $c > 0$, and let $(\rho_x)_{x \in S}$ be a near optimal sequence with average expected degree $\bar{\Delta} > 2$. Then for sufficiently large $n$ there is a threshold $c^*(\bar{\Delta})$ such that the random graph $G_S$ has the following property.

- If $c < c^*$, then $G_S$ has a matching with high probability.

# Existence of Thresholds

**Proposition**

Let $n = c \cdot m$, for constant $c > 0$, and let $(\rho_x)_{x \in S}$ be a near optimal sequence with average expected degree $\bar{\Delta} > 2$. Then for sufficiently large $n$ there is a threshold $c^*(\bar{\Delta})$ such that the random graph $G_S$ has the following property.

- If $c < c^*$, then $G_S$ has a matching with high probability.
- If $c > c^*$, then $G_S$ has no matching with high probability.

# Existence of Thresholds

**Proposition**

Let $n = c \cdot m$, for constant $c > 0$, and let $(\rho_x)_{x \in S}$ be a near optimal sequence with average expected degree $\bar{\Delta} > 2$. Then for sufficiently large $n$ there is a threshold $c^*(\bar{\Delta})$ such that the random graph $G_S$ has the following property.

- If $c < c^*$, then $G_S$ has a matching with high probability.
- If $c > c^*$, then $G_S$ has no matching with high probability.

This follows directly by the proofs for the thresholds of $k$-ary cuckoo hashing (with integral and non-integral $k = \bar{\Delta}$) using standard techniques on concentration bounds for nodes of certain degrees.

# Experimental Failure Rate Around the Threshold

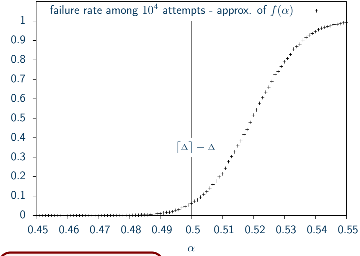
$(n, m, \bar{\Delta})$ is fixed, only two possible degrees $\lfloor \bar{\Delta} \rfloor$ and $\lceil \bar{\Delta} \rceil$

$\alpha$ is fraction of left nodes with degree $\lfloor \bar{\Delta} \rfloor$

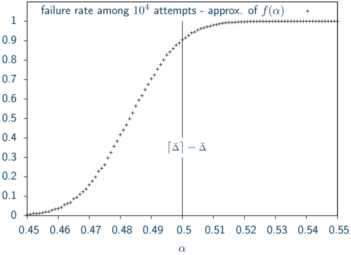
$f(\alpha)$ is rate of random graphs that have no matching — failure rate

$$c^*(\bar{\Delta}) = 0.957$$

$c = 0.956$




$c = 0.958$




**Conjecture**

# Experimental Failure Rate Around the Threshold

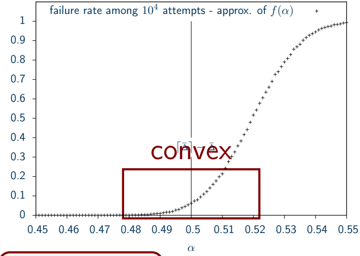
$(n, m, \bar{\Delta})$ is fixed, only two possible degrees $\lfloor \bar{\Delta} \rfloor$ and $\lceil \bar{\Delta} \rceil$

$\alpha$ is fraction of left nodes with degree $\lfloor \bar{\Delta} \rfloor$

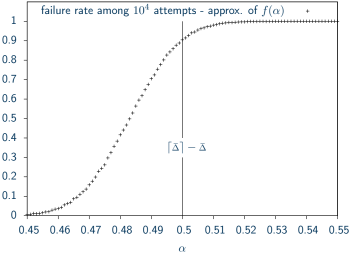
$f(\alpha)$ is rate of random graphs that have no matching — failure rate

$$c^*(\bar{\Delta}) = 0.957$$

$c = 0.956$




$c = 0.958$




**Conjecture**

If $c$ is to the left of $c^*(\bar{\Delta})$ then it is optimal to fix the degrees.

# Experimental Failure Rate Around the Threshold

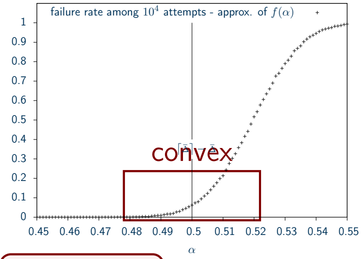
$(n, m, \bar{\Delta})$ is fixed, only two possible degrees $\lfloor \bar{\Delta} \rfloor$ and $\lceil \bar{\Delta} \rceil$

$\alpha$ is fraction of left nodes with degree $\lfloor \bar{\Delta} \rfloor$

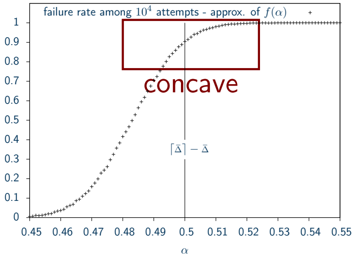
$f(\alpha)$ is rate of random graphs that have no matching — failure rate

$$c^*(\bar{\Delta}) = 0.957$$

$c = 0.956$




$c = 0.958$




**Conjecture**

If $c$ is to the left of $c^*(\bar{\Delta})$ then it is optimal to fix the degrees.
If $c$ is to the right of $c^*(\bar{\Delta})$ then it is optimal to use the Binomial distribution.

# Conclusion

# Random Bipartite Graphs

## Essence

# Random Bipartite Graphs

## Essence

1. Irregular degrees do not increase the probability of the existence of a matching.

# Random Bipartite Graphs

## Essence

1. Irregular degrees do not increase the probability of the existence of a matching.
2. Irregularity might improve the probability of the existence of a matching that can be calculated in linear time via a simple peeling algorithm (e.g., see erasure correcting codes).

# Random Bipartite Graphs

## Essence

1. Irregular degrees do not increase the probability of the existence of a matching.
2. Irregularity might improve the probability of the existence of a matching that can be calculated in linear time via a simple peeling algorithm (e.g., see erasure correcting codes).

## Open problem

Prove the conjecture.

Thank you!

# References (1)


**D, Goerdt, Mitzenmacher, Montanari, Pagh, R (2010).**
Tight Thresholds for Cuckoo Hashing via XORSAT.
In *Proc. 37th ICALP (1)*, LNCS, pages 213–225.

**Fountoulakis, N. and Panagiotou, K. (2010).**
Orientability of Random Hypergraphs and the Power of Multiple Choices.
In *Proc. 37th ICALP (1)*, LNCS, pages 348–359.

**Frieze, A. M. and Melsted, P. (2009).**
Maximum Matchings in Random Bipartite Graphs and the Space Utilization of Cuckoo Hashtables.
*CoRR*, arXiv:0910.5535.

**Luby, M. (2002).**
LT Codes.
In *Proc. 43rd FOCS*, pages 271–. IEEE Computer Society.

# References (2)


**Luby, Mitzenmacher, Shokrollahi, Spielman (2001).**
Efficient Erasure Correcting Codes.
*IEEE Transactions on Information Theory*, 47(2):569–584.

**Maymounkov, P. (2002).**
Online codes (Extended Abstract).
Technical Report TR2002-833, New York University.

**Shokrollahi, A. (2006).**
Raptor Codes.
*IEEE Transactions on Information Theory*, 52(6):2551–2567.