Reconstructing Random Pictures

Corrine Yap

Georgia Tech

Joint work with Bhargav Narayanan

AMS Central Sectional

October 7–8, 2023

Background

Reconstruction Problem

Given a discrete structure, can we uniquely reconstruct it from the list of its substructures of a fixed size?

Most famous example: graphs—Vertex and Edge Reconstruction Conjectures (Kelly, Ulam 1957, Harary 1964)

Mossel–Ross '18

What about "shotgun assembly?" (motivated by shotgun sequencing of DNA)

Background

Reconstruction Problem

Given a discrete structure, can we uniquely reconstruct it from the list of its substructures of a fixed size?

Most famous example: graphs—Vertex and Edge Reconstruction Conjectures (Kelly, Ulam 1957, Harary 1964)

Mossel–Ross '18

What about "shotgun assembly?" (motivated by shotgun sequencing of DNA)

Background

Reconstruction Problem

Given a discrete structure, can we uniquely reconstruct it from the list of its substructures of a fixed size?

Most famous example: graphs—Vertex and Edge Reconstruction Conjectures (Kelly, Ulam 1957, Harary 1964)

Mossel–Ross '18

What about "shotgun assembly?" (motivated by shotgun sequencing of DNA)

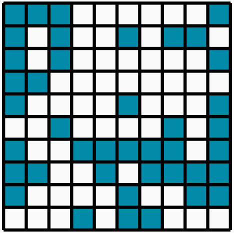
Random Pictures

Today: Let P_n be a random picture, i.e. an $n \times n$ grid with $\{0, 1\}$ entries chosen uniformly at random. Let $\mathcal{D}$ be the deck of its $k \times k$ subgrids.

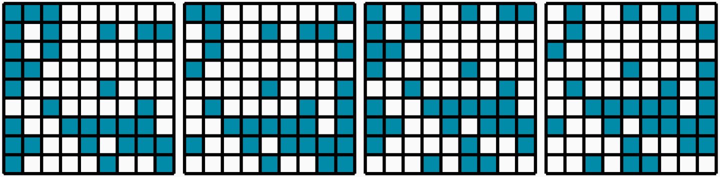
Question

For what $k = k(n)$ is P_n reconstructible from $\mathcal{D}$ with high probability?

Example

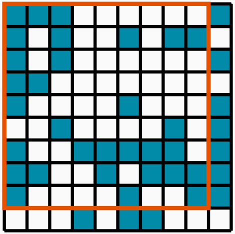


A 10 × 10 picture

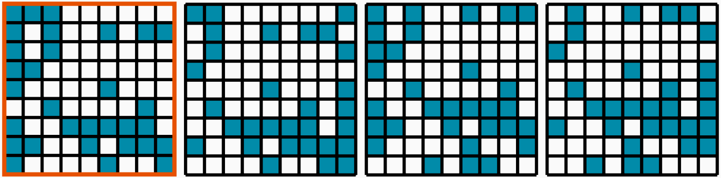


Deck of 9 × 9 subgrids

Example

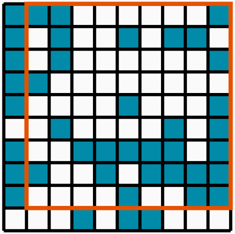


A 10 × 10 picture

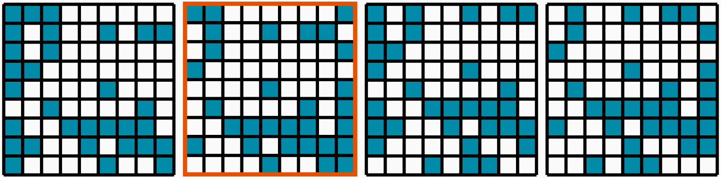


Deck of 9 × 9 subgrids

Example

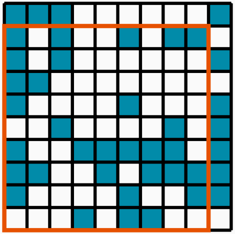


A 10 × 10 picture

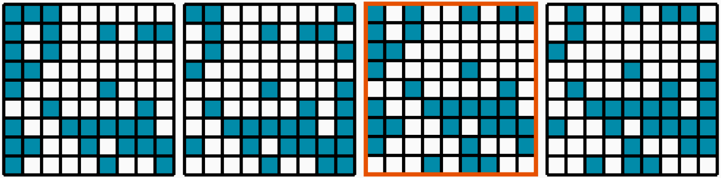


Deck of 9 × 9 subgrids

Example

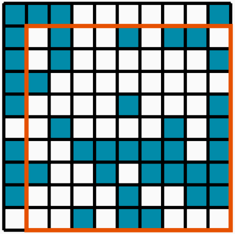


A 10 × 10 picture

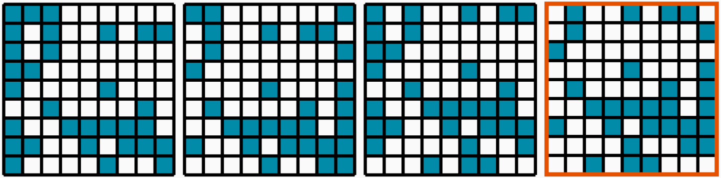


Deck of 9 × 9 subgrids

Example

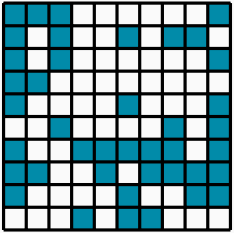


A 10 × 10 picture

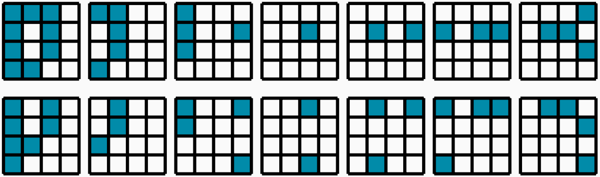


Deck of 9 × 9 subgrids

Example



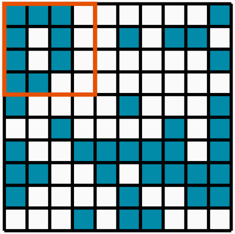
A 10×10 picture



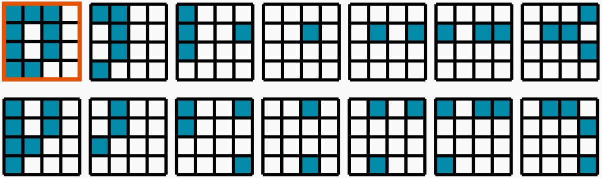
...

Deck of 4×4 subgrids

Example



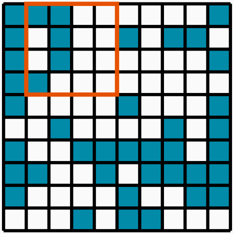
A 10×10 picture



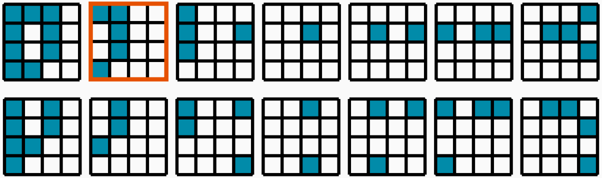
...

Deck of 4×4 subgrids

Example



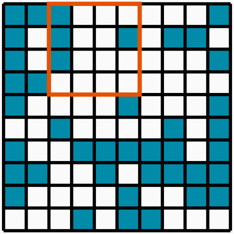
A 10×10 picture



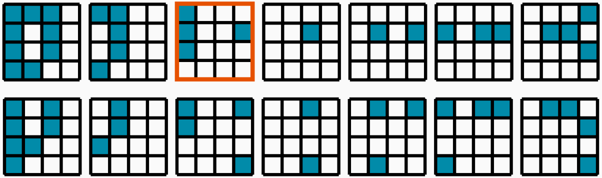
...

Deck of 4×4 subgrids

Example



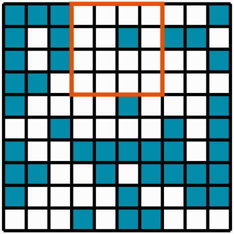
A 10×10 picture



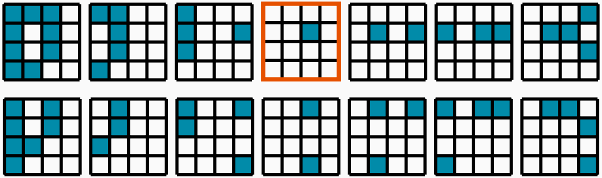
...

Deck of 4×4 subgrids

Example



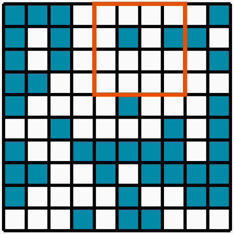
A 10×10 picture



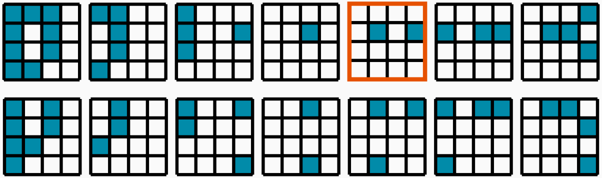
...

Deck of 4×4 subgrids

Example



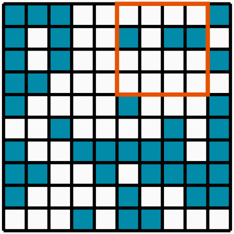
A 10×10 picture



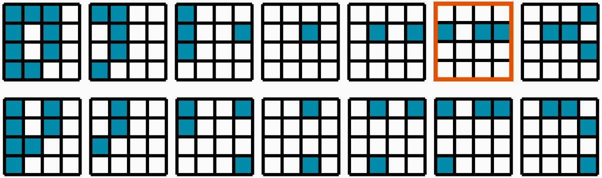
. . .

Deck of 4×4 subgrids

Example



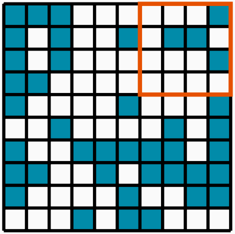
A 10×10 picture



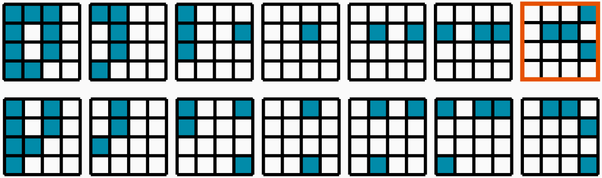
...

Deck of 4×4 subgrids

Example



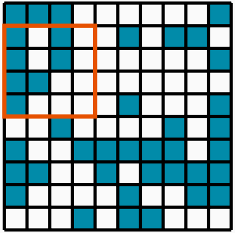
A 10 × 10 picture



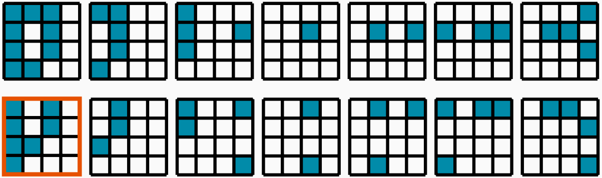
. . .

Deck of 4 × 4 subgrids

Example



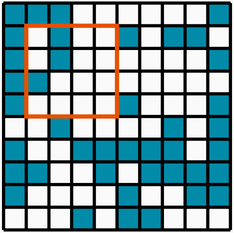
A 10×10 picture



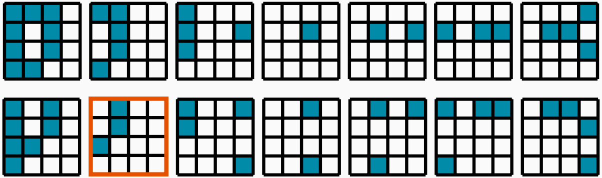
...

Deck of 4×4 subgrids

Example



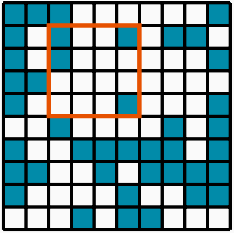
A 10×10 picture



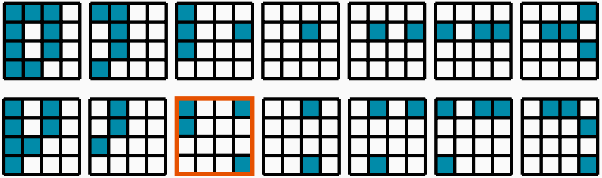
...

Deck of 4×4 subgrids

Example



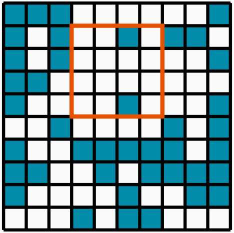
A 10 × 10 picture



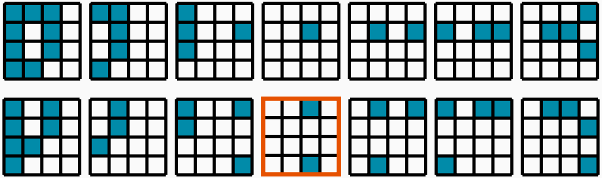
. . .

Deck of 4 × 4 subgrids

Example



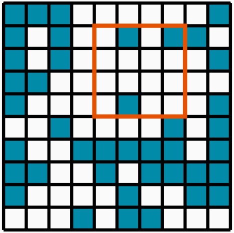
A 10×10 picture



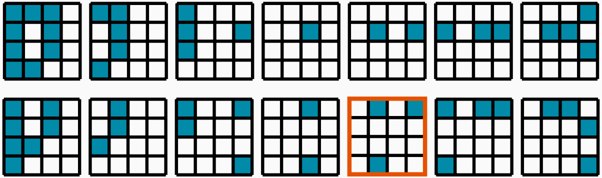
...

Deck of 4×4 subgrids

Example



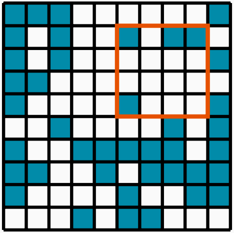
A 10 × 10 picture



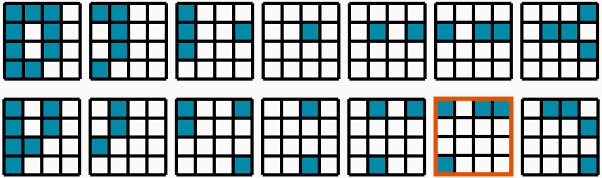
. . .

Deck of 4 × 4 subgrids

Example



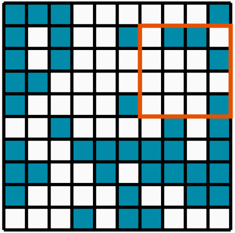
A 10×10 picture



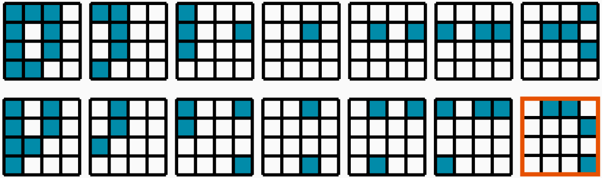
. . .

Deck of 4×4 subgrids

Example



A 10×10 picture



...

Deck of 4×4 subgrids

Main Theorem

Let $R(n,k)$ be the event that P_n is reconstructible from its k-deck.

Narayanan-Y. '23+

There exists $k_c(n)$ such that as $n \to \infty$,

$$\mathrm{Prob}[R(n,k)] \to \begin{cases} 0 & \text{if } k < k_c(n) \\ 1 & \text{if } k > k_c(n) \end{cases}$$

Moreover, $k_c(n)$ takes one of two values: $\lfloor \sqrt{2\log_2 n} \rfloor, \lceil \sqrt{2\log_2 n} \rceil$.

Main Theorem

Let $R(n,k)$ be the event that P_n is reconstructible from its k-deck.

Narayanan-Y. '23+

There exists $k_c(n)$ such that as $n \to \infty$,

$$\mathrm{Prob}[R(n,k)] \to \begin{cases} 0 & \text{if } k < k_c(n) \\ 1 & \text{if } k > k_c(n) \end{cases}$$

Moreover, $k_c(n)$ takes one of two values: $\lfloor \sqrt{2\log_2 n} \rfloor, \lceil \sqrt{2\log_2 n} \rceil$.

Proof of the 0-Statement: If $k < k_c(n)$, then $n^2 2^{-k^2} \to \infty$ as $n \to \infty$.

Counting argument; bound the number of reconstructible pictures by the number of k-decks.

Main Theorem

Let $R(n,k)$ be the event that P_n is reconstructible from its k-deck.

Narayanan-Y. '23+

There exists $k_c(n)$ such that as $n \to \infty$,

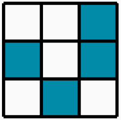
$$\mathrm{Prob}[R(n,k)] \to \begin{cases} 0 & \text{if } k < k_c(n) \\ 1 & \text{if } k > k_c(n) \end{cases}$$

Moreover, $k_c(n)$ takes one of two values: $\lfloor \sqrt{2\log_2 n} \rfloor, \lceil \sqrt{2\log_2 n} \rceil$.

Proof of the 1-Statement: If $k > k_c(n)$, then $n^2 k 2^{-k^2+k} \to 0$. Our goal is to give an algorithm for reconstructing P_n from its deck and prove that the probability of failure tends to 0.

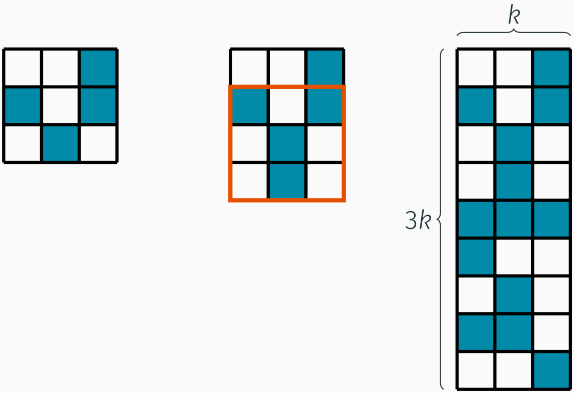
Reconstruction Algorithm

Step 0: Randomly order the deck $\mathcal{D}$ and begin with the first deck element.



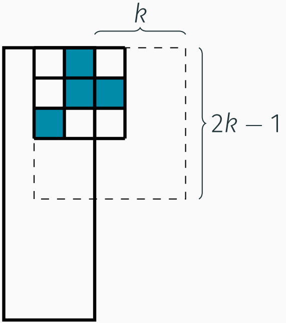
Reconstruction Algorithm

Step 1: Extend downward to $3k$ rows by placing the first deck element that fits.



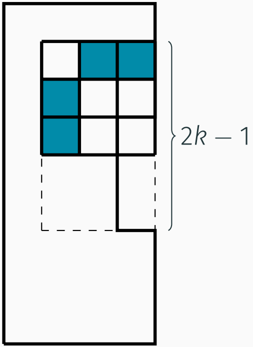
Reconstruction Algorithm

Step 2: Extend to the right one column at a time, first at each of the corners



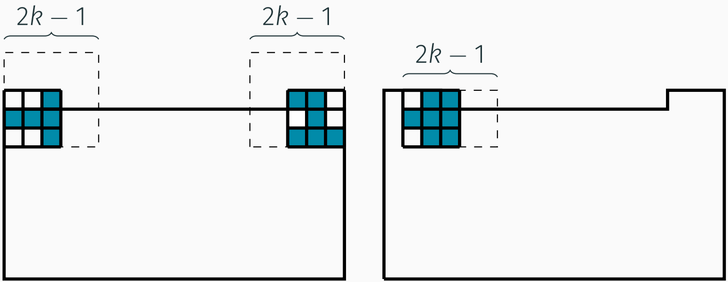
Reconstruction Algorithm

Step 2: Extend to the right one column at a time, first at each of the corners then internally. Repeat to the right and left until n columns.

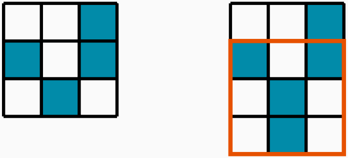


Reconstruction Algorithm

Step 3: Extend upward one row at a time, then downward until n rows.



Analysis: Naive Extensions



Observe that for each naive extension,

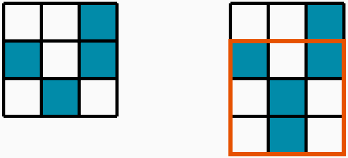
$$\mathrm{Prob}[\text{mistake}] \leq n^2 2^{-k^2+k}$$

So by union bound,

$$\mathrm{Prob}[\text{there is a mistake in the first step}] \leq 3kn^2 2^{-k^2+k}$$

which tends to 0 by our assumption. However, we cannot afford to do naive extensions for the entire grid. This is why we introduce the corner and internal extensions.

Analysis: Naive Extensions



Observe that for each naive extension,

$$\mathrm{Prob}[\text{mistake}] \leq n^2 2^{-k^2+k}$$

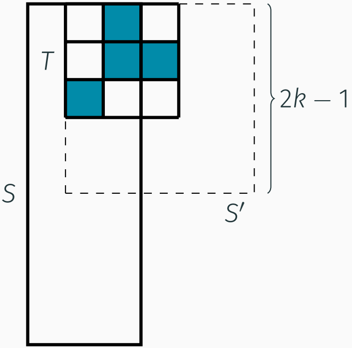
So by union bound,

$$\mathrm{Prob}[\text{there is a mistake in the first step}] \leq 3kn^2 2^{-k^2+k}$$

which tends to 0 by our assumption. However, we cannot afford to do naive extensions for the entire grid. This is why we introduce the corner and internal extensions.

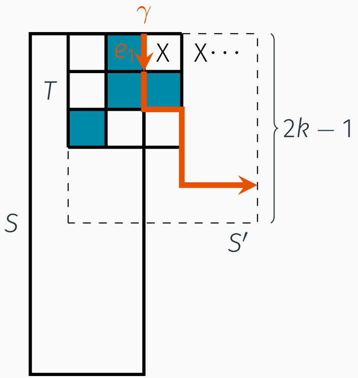
Analysis: Corner Extensions

Suppose we have correctly reconstructed S and are extending to the right. Before placing a corner subgrid T, we check to see if it can be extended to a $(2k-1) \times (2k-1)$ subgrid S' using deck elements.



Analysis: Corner Extensions

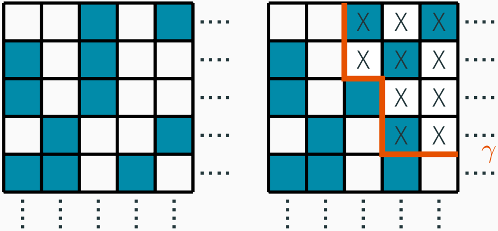
Suppose we have correctly reconstructed S and are extending to the right. Before placing a corner subgrid T, we check to see if it can be extended to a $(2k-1) \times (2k-1)$ subgrid S' using deck elements.



A k-grid is *bad* if it is incorrect with respect to P_n. We mark bad k-grids, e.g. in the upper-right corner.

An interface path is a path separating the good and bad entries.

Analysis: Interface Paths



We compute probabilities associated with the interface paths. For example,

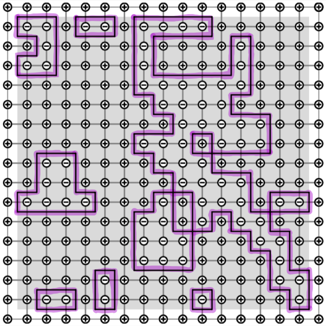
$$\text{Prob}[\text{first step}] \leq n^2 2^{-k^2+k}$$

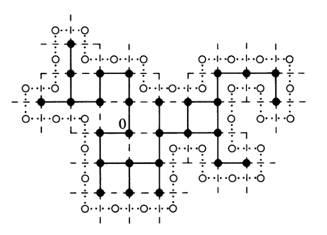
but

$$\text{Prob}[\text{second step} \mid \text{first step}] \leq n^2 2^{-k^2+1} + 2(4k^2)(2^{-k+1})$$

Digression

The technique of computing a first moment along a path/contour originated with Peierls in a proof of phase coexistence for the Ising model on $\mathbb{Z}^d$ and is often used in percolation.





Images from Friedli-Velenik, *Statistical Mechanics of Lattice Systems* and Grimmett, *Percolation*

Further Directions

- Demidovich–Panichkin–Zhukovskii use a variation of our techniques to give 2-point concentration for dimensions $d \geq 2$ and colors $r \geq 2$
- Sharp threshold?
- DPZ also connects their results to reconstruction of uniform r-colorings of $G(n, 1/2)$ from k-decks (neighborhoods of radius k), but there is a gap from $\sqrt{\log_2(n)}$ to $\log_2 n$.
- More variants: non-square, p-biased, noisy, correlated...

Further Directions

- Demidovich–Panichkin–Zhukovskii use a variation of our techniques to give 2-point concentration for dimensions $d \geq 2$ and colors $r \geq 2$
- Sharp threshold?
- DPZ also connects their results to reconstruction of uniform r-colorings of $G(n, 1/2)$ from k-decks (neighborhoods of radius k), but there is a gap from $\sqrt{\log_2(n)}$ to $\log_2 n$.
- More variants: non-square, p-biased, noisy, correlated...

Further Directions

- Demidovich–Panichkin–Zhukovskii use a variation of our techniques to give 2-point concentration for dimensions $d \geq 2$ and colors $r \geq 2$
- Sharp threshold?
- DPZ also connects their results to reconstruction of uniform r-colorings of $G(n, 1/2)$ from k-decks (neighborhoods of radius k), but there is a gap from $\sqrt{\log_2(n)}$ to $\log_2 n$.
- More variants: non-square, p-biased, noisy, correlated...

Further Directions

- Demidovich–Panichkin–Zhukovskii use a variation of our techniques to give 2-point concentration for dimensions $d \geq 2$ and colors $r \geq 2$
- Sharp threshold?
- DPZ also connects their results to reconstruction of uniform r-colorings of $G(n, 1/2)$ from k-decks (neighborhoods of radius k), but there is a gap from $\sqrt{\log_2(n)}$ to $\log_2 n$.
- More variants: non-square, p-biased, noisy, correlated...

Thank you!