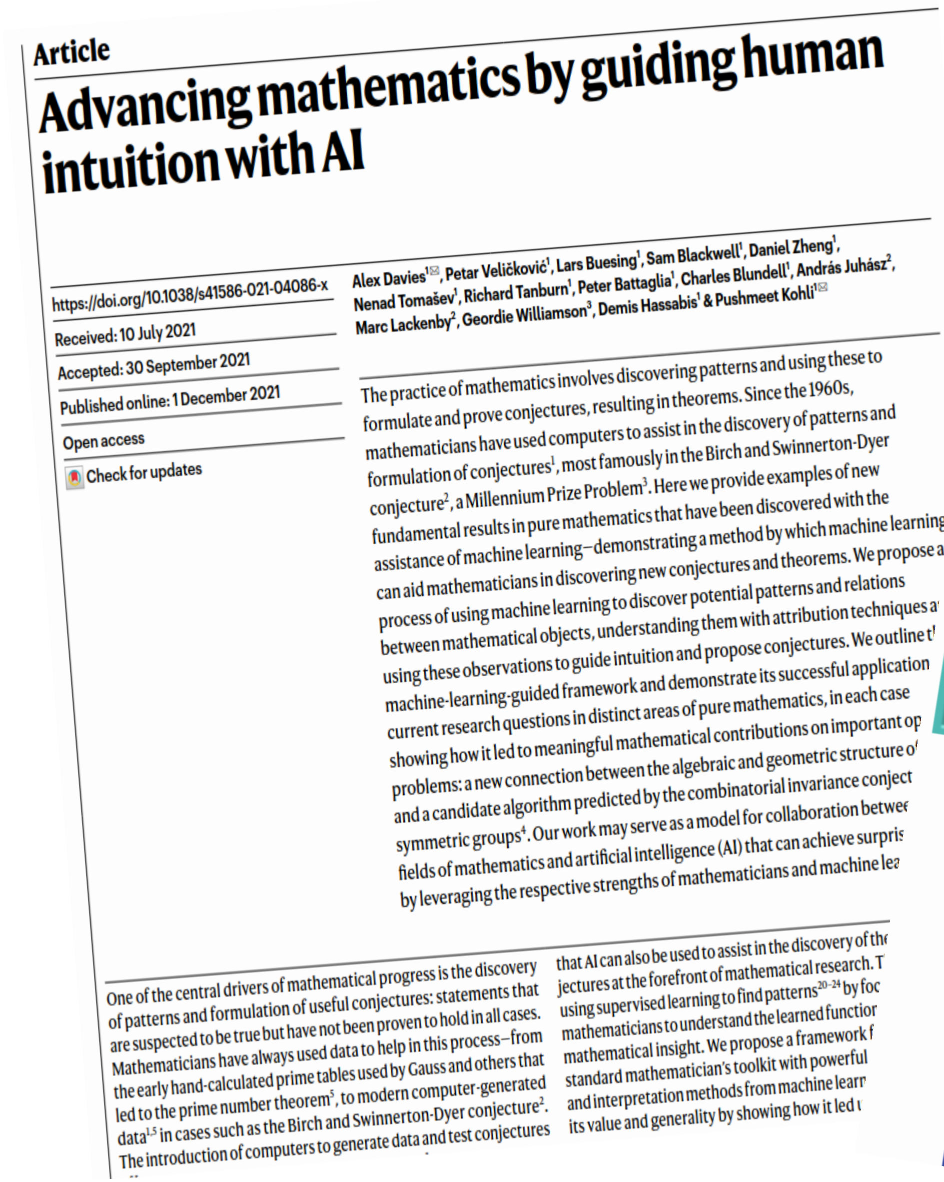
Mittagsseminar

Constructions in combinatorics via neuronal networks

(Wagner, 2021)

Talk by: Sandro Roch



Article

Advancing mathematics by guiding human intuition with AI

https://doi.org/10.1038/s41586-021-04086-x

Received: 10 July 2021

Accepted: 30 September 2021

Published online: 1 December 2021

Open access

Check for updates

Alex Davies[1]✉, Petar Veličković[1], Lars Buesing[1], Sam Blackwell[1], Daniel Zheng[1], Nenad Tomašev[1], Richard Tanburn[1], Peter Battaglia[1], Charles Blundell[1], András Juhász[2], Marc Lackenby[2], Geordie Williamson[3], Demis Hassabis[1] & Pushmeet Kohli[1]✉

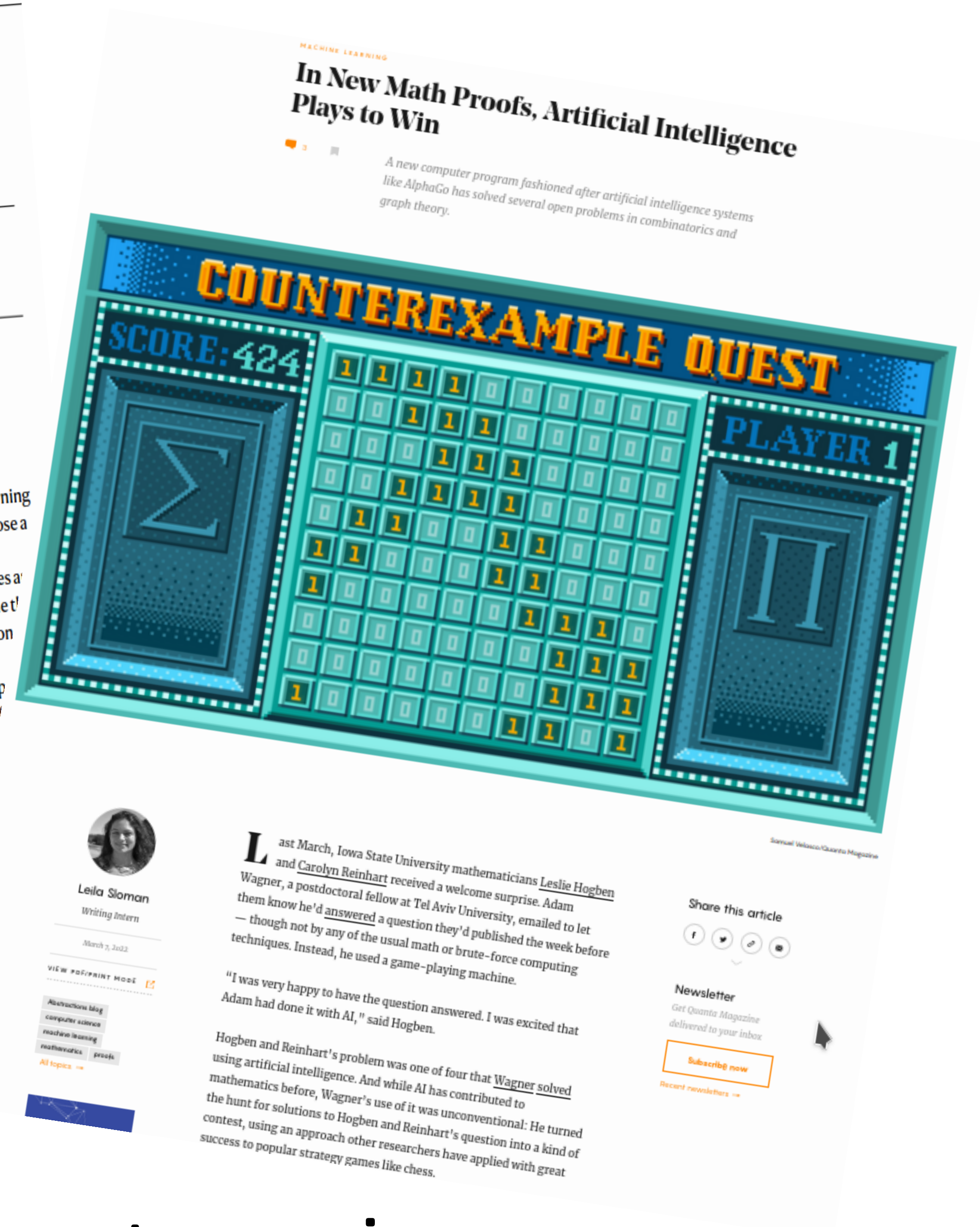
The practice of mathematics involves discovering patterns and using these to formulate and prove conjectures, resulting in theorems. Since the 1960s, mathematicians have used computers to assist in the discovery of patterns and formulation of conjectures[1], most famously in the Birch and Swinnerton-Dyer conjecture[2], a Millennium Prize Problem[3]. Here we provide examples of new fundamental results in pure mathematics that have been discovered with the assistance of machine learning—demonstrating a method by which machine learning can aid mathematicians in discovering new conjectures and theorems. We propose a process of using machine learning to discover potential patterns and relations between mathematical objects, understanding them with attribution techniques a using these observations to guide intuition and propose conjectures. We outline t machine-learning-guided framework and demonstrate its successful application current research questions in distinct areas of pure mathematics, in each case showing how it led to meaningful mathematical contributions on important op problems: a new connection between the algebraic and geometric structure o and a candidate algorithm predicted by the combinatorial invariance conject symmetric groups[4]. Our work may serve as a model for collaboration betwe fields of mathematics and artificial intelligence (AI) that can achieve surpris by leveraging the respective strengths of mathematicians and machine lea

One of the central drivers of mathematical progress is the discovery of patterns and formulation of useful conjectures: statements that are suspected to be true but have not been proven to hold in all cases. Mathematicians have always used data to help in this process—from the early hand-calculated prime tables used by Gauss and others that led to the prime number theorem[5], to modern computer-generated data[1,5] in cases such as the Birch and Swinnerton-Dyer conjecture[2]. The introduction of computers to generate data and test conjectures

that AI can also be used to assist in the discovery of th jectures at the forefront of mathematical research. T using supervised learning to find patterns[20–24] by foc mathematicians to understand the learned functior mathematical insight. We propose a framework f standard mathematician's toolkit with powerful and interpretation methods from machine learr its value and generality by showing how it led t

nature

Article citing Wagner



quantamagazine.com

Article dedicated to Wagner's work

Generating structures using reinforcement learning

Generating structures using reinforcement learning

- Encode instances by fixed length word w over fin. alphabet

  Example: Graph on n vertices as $w \in \{0,1\}^{\binom{n}{2}}$

Generating structures using reinforcement learning

- Encode instances by fixed length word w over fin. alphabet

  Example: Graph on n vertices as $w \in \{0,1\}^{\binom{n}{2}}$

- Start with empty word

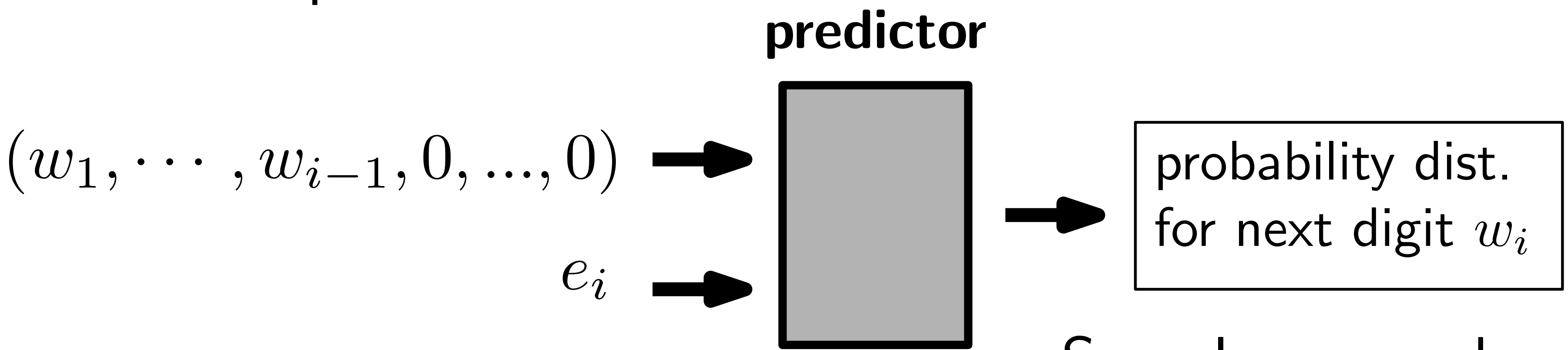
Generating structures using reinforcement learning

- Encode instances by fixed length word w over fin. alphabet

  Example: Graph on n vertices as $w \in \{0,1\}^{\binom{n}{2}}$

- Start with empty word

- In i-th step:



Sample w_i randomly!

Generating structures using reinforcement learning

Generating structures using reinforcement learning

Deep cross-entropy method:

Generating structures using reinforcement learning

Deep cross-entropy method:

- Generate $N > 0$ instances

Generating structures using reinforcement learning

Deep cross-entropy method:

- Generate $N > 0$ instances
- For each instance: Evaluate score function
  Ex.: *„How close is instance to conjectured bound?“*

Generating structures using reinforcement learning

Deep cross-entropy method:

- Generate $N > 0$ instances
- For each instance: Evaluate score function
  Ex.: *„How close is instance to conjectured bound?“*
- For top y percentage of instances:

  Training: Fit predictor on pairs

$$((w_1, ..., w_{i-1}, 0, ..., 0),\ \ e_i) \ \ \rightarrow \ \ e_{w_i}$$

Generating structures using reinforcement learning

Deep cross-entropy method:

- Generate $N > 0$ instances
- For each instance: Evaluate score function
  Ex.: „*How close is instance to conjectured bound?*“
- For top y percentage of instances:

  Training: Fit predictor on pairs

  $$((w_1, ..., w_{i-1}, 0, ..., 0),\ \ e_i) \ \ \rightarrow \ \ e_{w_i}$$

- Keep top $x < y$ percentage of instances
  for next iteration(s)

Generating structures using reinforcement learning

Generating structures using reinforcement learning

Architecture of predictor

Generating structures using reinforcement learning

Architecture of predictor

- Neuronal network with three hidden layers:
  dense layers with 128 / 64 / 4 nodes
  activation function: ReLU

Generating structures using reinforcement learning

Architecture of predictor

- Neuronal network with three hidden layers:
  dense layers with 128 / 64 / 4 nodes
  activation function: ReLU
- Output layer: sigmoid (binary case)

Generating structures using reinforcement learning

Architecture of predictor

- Neuronal network with three hidden layers:
  dense layers with 128 / 64 / 4 nodes
  activation function: ReLU
- Output layer: sigmoid (binary case)
- Loss function: Cross entropy

Generating structures using reinforcement learning

Architecture of predictor

- Neuronal network with three hidden layers:
  dense layers with 128 / 64 / 4 nodes
  activation function: ReLU
- Output layer: sigmoid (binary case)
- Loss function: Cross entropy
- Optimizer: SGD

Example

M. Aouchiche and P. Hansen, *A survey of automated conjectures in spectral graph theory*, 2010:

Conjecture: G connected graph, $n \geq 3$, largest eigenvalue λ_1, matching number μ. Then:

$$\lambda_1 + \mu \geq \sqrt{n-1} + 1$$

Example

M. Aouchiche and P. Hansen, *A survey of automated conjectures in spectral graph theory*, 2010:

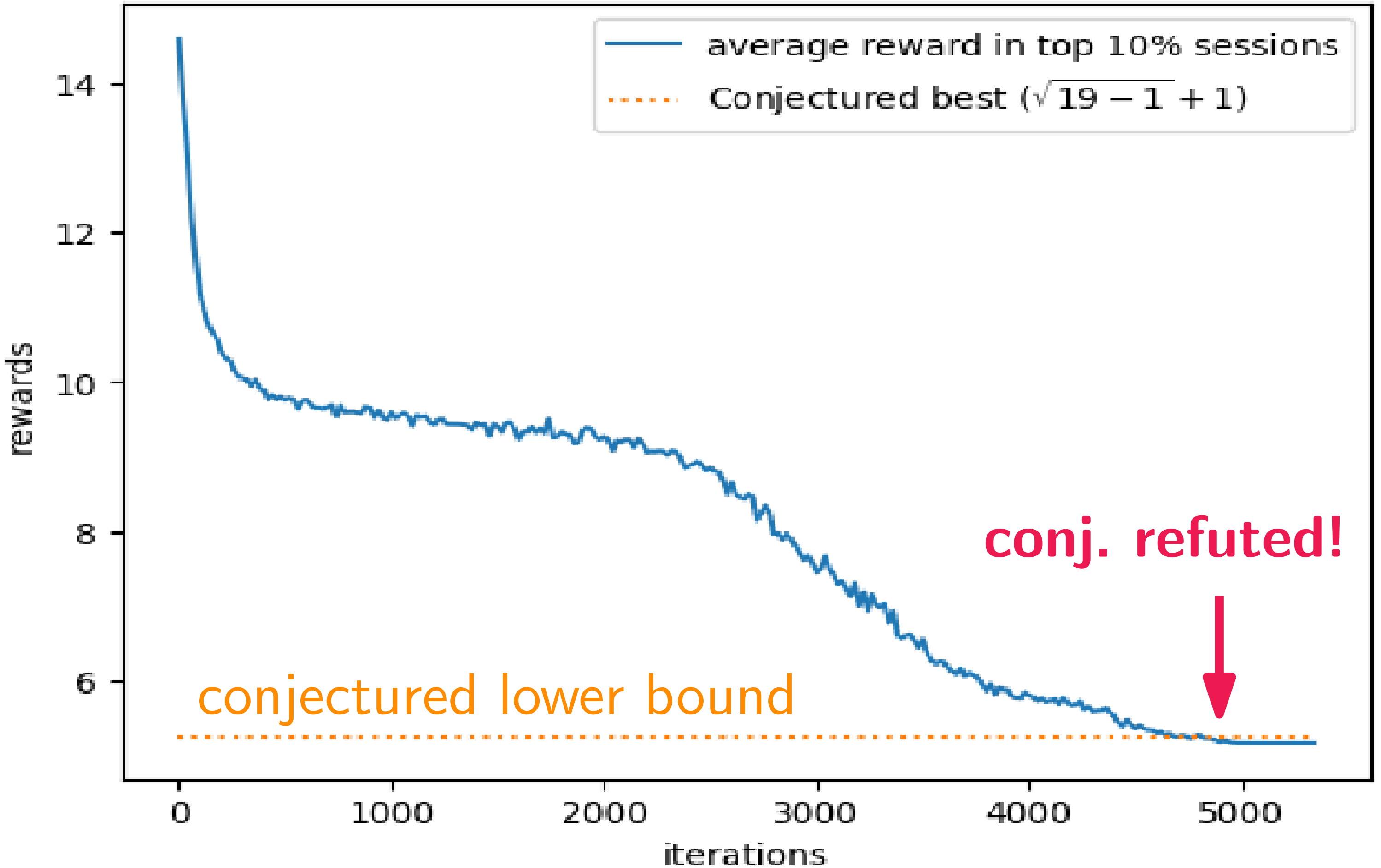
Conjecture: G connected graph, $n \geq 3$, largest eigenvalue λ_1, matching number μ. Then:

$$\lambda_1 + \mu \geq \sqrt{n-1} + 1$$

Apply cross-entropy method:
Fix n, minimize score function $\lambda_1 + \mu$

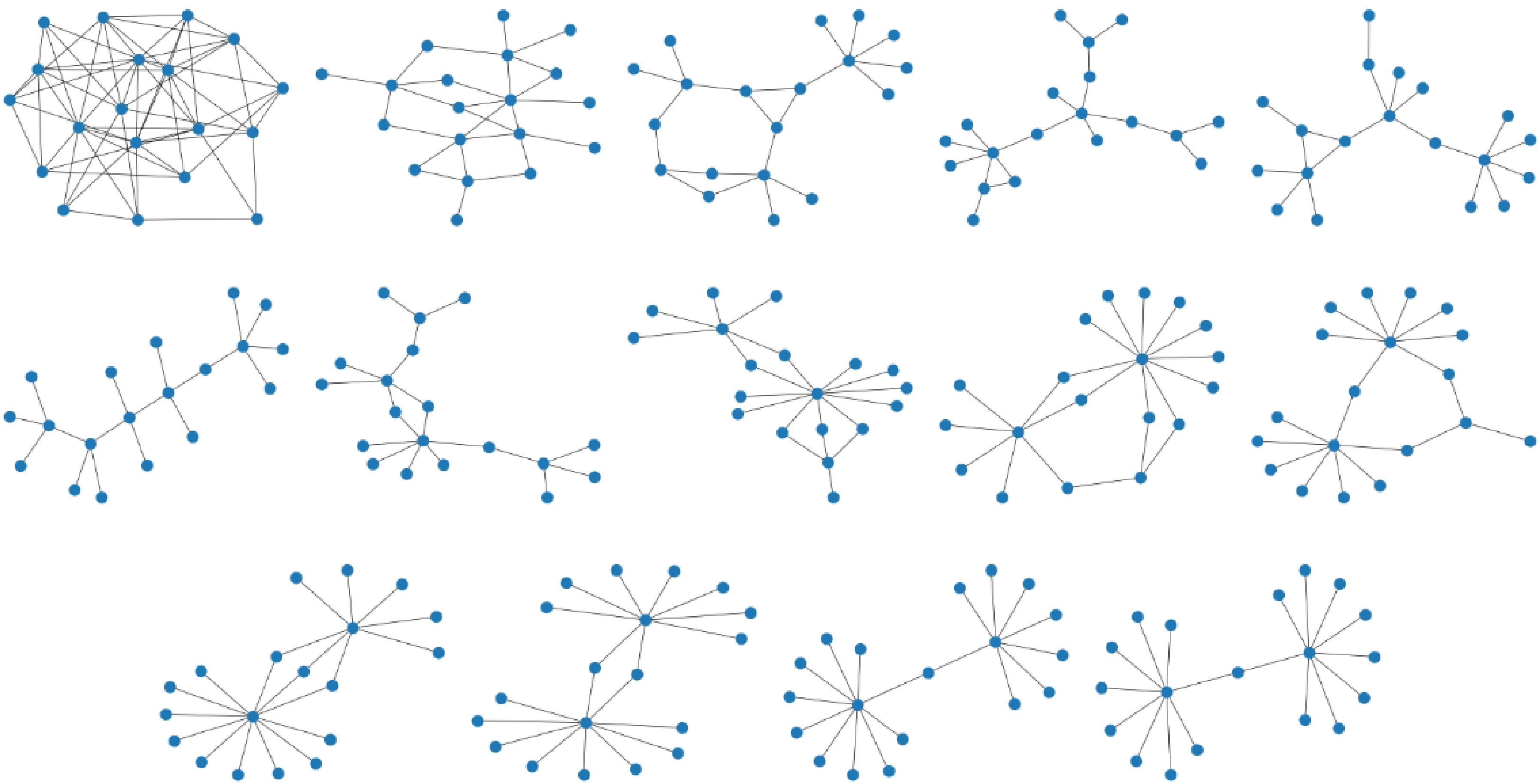
Example

For $n = 19$, average score after many iterations:



Example

Evolution of best scoring instance:



conterexample

Pattern avoiding 0-1-matrices

Pattern avoiding 0-1-matrices

Definition: 0-1-**matrix patterns**

The 0-1-matrix $A \in \{0,1\}^{r \times s}$ *contains* 0-1-matrix $P \in \{0,1\}^{k \times l}$, if there exists a submatrix $D \in \{0,1\}^{k \times l}$ with $P \leq D$.

Pattern avoiding 0-1-matrices

Definition: 0-1-matrix patterns

The 0-1-matrix $A \in \{0,1\}^{r \times s}$ *contains* 0-1-matrix $P \in \{0,1\}^{k \times l}$, if there exists a submatrix $D \in \{0,1\}^{k \times l}$ with $P \leq D$.

Otherwise, A *avoids* P.

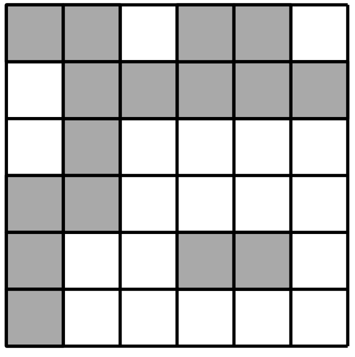
Pattern avoiding 0-1-matrices

Definition: 0-1-matrix patterns

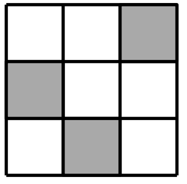
The 0-1-matrix $A \in \{0,1\}^{r \times s}$ *contains* 0-1-matrix $P \in \{0,1\}^{k \times l}$, if there exists a submatrix $D \in \{0,1\}^{k \times l}$ with $P \leq D$.

Otherwise, A *avoids* P.

Example:



contains



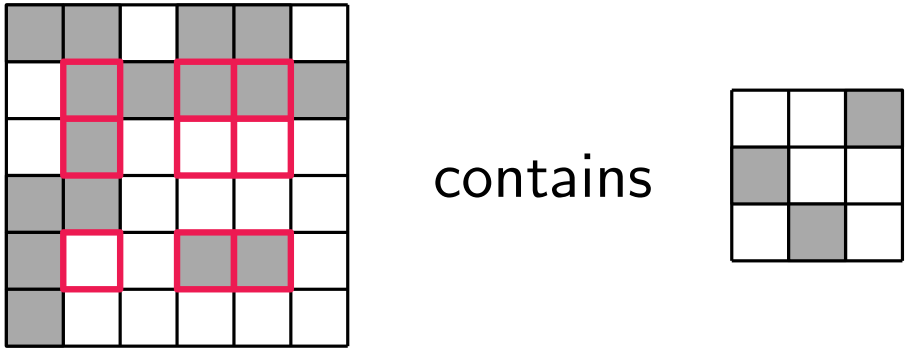
Pattern avoiding 0-1-matrices

Definition: 0-1-matrix patterns

The 0-1-matrix $A \in \{0,1\}^{r \times s}$ *contains* 0-1-matrix $P \in \{0,1\}^{k \times l}$, if there exists a submatrix $D \in \{0,1\}^{k \times l}$ with $P \leq D$.

Otherwise, A *avoids* P.

Example:



Pattern avoiding 0-1-matrices

Pattern avoiding 0-1-matrices

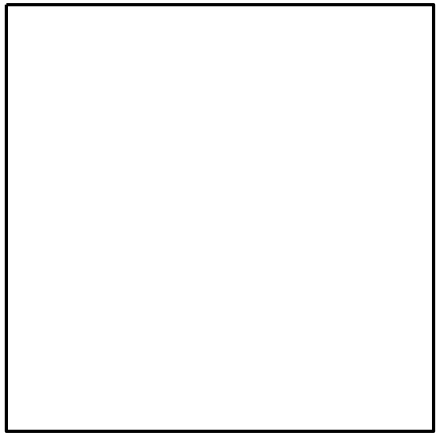
- $P_\pi := [\delta_{j,\pi(i)}]_{i,j} \in \{0,1\}^{n\times n}$ permutation matrix of $\pi \in S_n$

Pattern avoiding 0-1-matrices

- $P_\pi := [\delta_{j,\pi(i)}]_{i,j} \in \{0,1\}^{n\times n}$ permutation matrix of $\pi \in S_n$
- **Def:** 0-1-matrix M *contains* (resp. *avoids*) $\pi \in S_k$,
  if M avoids (resp. contains) P_π.

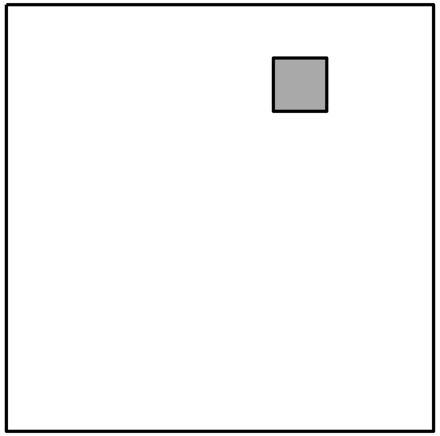
Pattern avoiding 0-1-matrices

- $P_\pi := [\delta_{j,\pi(i)}]_{i,j} \in \{0,1\}^{n \times n}$ permutation matrix of $\pi \in S_n$
- **Def:** 0-1-matrix M *contains* (resp. *avoids*) $\pi \in S_k$,
  if M avoids (resp. contains) P_π.
- **Example:** 312-avoiding 0-1-matrices:



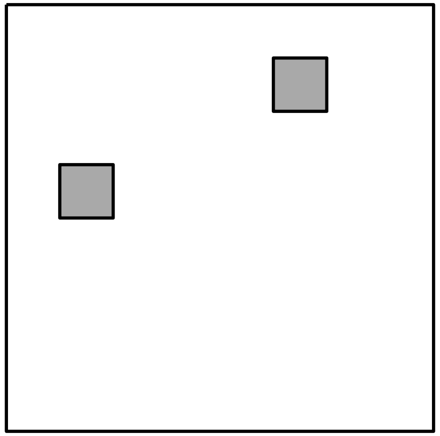
Pattern avoiding 0-1-matrices

- $P_\pi := [\delta_{j,\pi(i)}]_{i,j} \in \{0,1\}^{n\times n}$ permutation matrix of $\pi \in S_n$
- **Def:** 0-1-matrix M *contains* (resp. *avoids*) $\pi \in S_k$,
  if M avoids (resp. contains) P_π.
- **Example:** 312-avoiding 0-1-matrices:



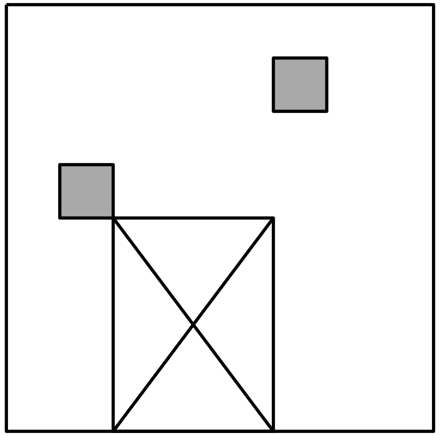
Pattern avoiding 0-1-matrices

- $P_\pi := [\delta_{j,\pi(i)}]_{i,j} \in \{0,1\}^{n\times n}$ permutation matrix of $\pi \in S_n$
- **Def:** 0-1-matrix M *contains* (resp. *avoids*) $\pi \in S_k$,
  if M avoids (resp. contains) P_π.
- **Example:** 312-avoiding 0-1-matrices:



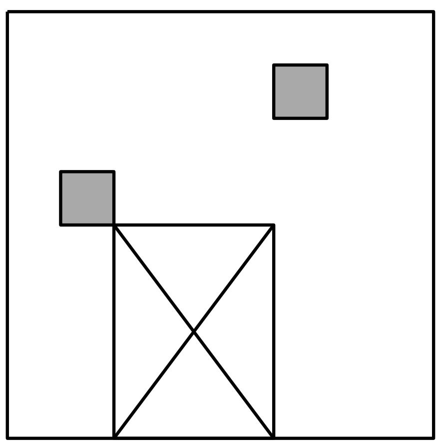
Pattern avoiding 0-1-matrices

- $P_\pi := [\delta_{j,\pi(i)}]_{i,j} \in \{0,1\}^{n\times n}$ permutation matrix of $\pi \in S_n$
- **Def:** 0-1-matrix M *contains* (resp. *avoids*) $\pi \in S_k$,
  if M avoids (resp. contains) P_π.
- **Example:** 312-avoiding 0-1-matrices:



Pattern avoiding 0-1-matrices

- $P_\pi := [\delta_{j,\pi(i)}]_{i,j} \in \{0,1\}^{n\times n}$ permutation matrix of $\pi \in S_n$
- **Def:** 0-1-matrix M *contains* (resp. *avoids*) $\pi \in S_k$,
  if M avoids (resp. contains) P_π.
- **Example:** 312-avoiding 0-1-matrices:



- **Obs.:** P_π avoids P_σ iff π avoids σ as permutation pattern.

Pattern avoiding 0-1-matrices

Conjecture (Füredi & Hajnal, 1992):
Let $\pi \in S_k$. The number of 1-entries in matrices $M \in \{0,1\}^{n \times n}$ avoiding π is bounded by $O(n)$.

Pattern avoiding 0-1-matrices

Conjecture (Füredi & Hajnal, 1992):
Let $\pi \in S_k$. The number of 1-entries in matrices $M \in \{0,1\}^{n \times n}$ avoiding π is bounded by $O(n)$.

Theorem (Marcus & Tardos, 2004):
Füredi-Hajnal-conjecture is true.

Proof: Later

Permanent in pattern avoiding 0-1-matrices

Permanent in pattern avoiding 0-1-matrices

Motivation: If $A \in \{0,1\}^{n \times n}$ avoids π, then the permanent

$$\mathrm{per}(A) := \sum_{\sigma \in S_n} \prod_{i=1}^{n} a_{i,\sigma(i)}$$

counts (π-avoiding) permutations S_n contained in A.

Permanent in pattern avoiding 0-1-matrices

Motivation: If $A \in \{0,1\}^{n\times n}$ avoids π, then the permanent

$$\mathrm{per}(A) := \sum_{\sigma \in S_n} \prod_{i=1}^{n} a_{i,\sigma(i)}$$

counts (π-avoiding) permutations S_n contained in A.

Question (Brualdi & Cao, 2020):
Given $n \in \mathbb{N}$ and $\pi \in S_k$, what is the value of

$$f_\pi(n) := \max\left\{\mathrm{per}(A) : A \in \{0,1\}^{n\times n}, A \text{ avoids } \pi\right\}$$

Permanent in pattern avoiding 0-1-matrices

Motivation: If $A \in \{0,1\}^{n \times n}$ avoids π, then the permanent

$$\mathrm{per}(A) := \sum_{\sigma \in S_n} \prod_{i=1}^{n} a_{i,\sigma(i)}$$

counts (π-avoiding) permutations S_n contained in A.

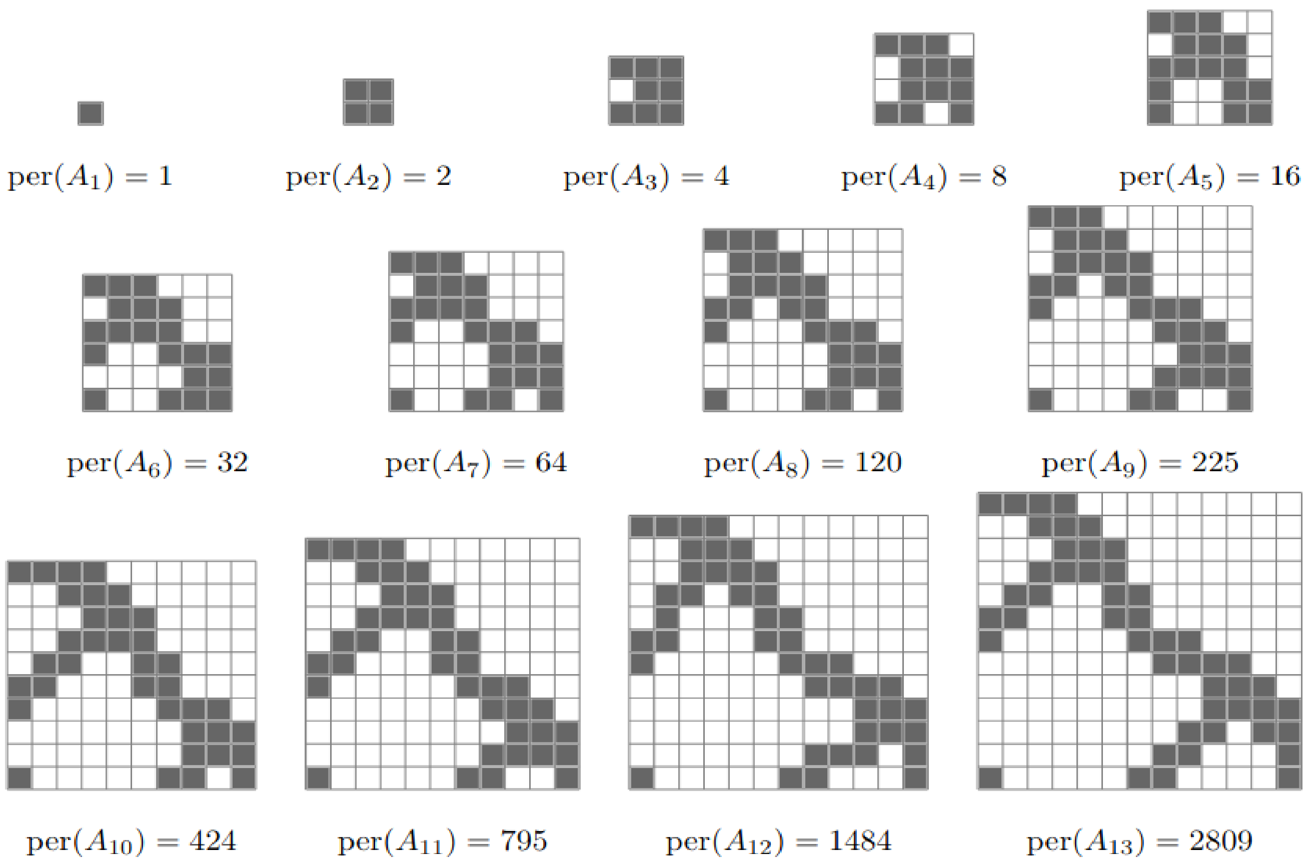
Question (Brualdi & Cao, 2020):
Given $n \in \mathbb{N}$ and $\pi \in S_k$, what is the value of

$$f_\pi(n) := \max\left\{\mathrm{per}(A) : A \in \{0,1\}^{n \times n}, A \text{ avoids } \pi\right\}$$

Here: Wagner finds bounds on f_{312}.

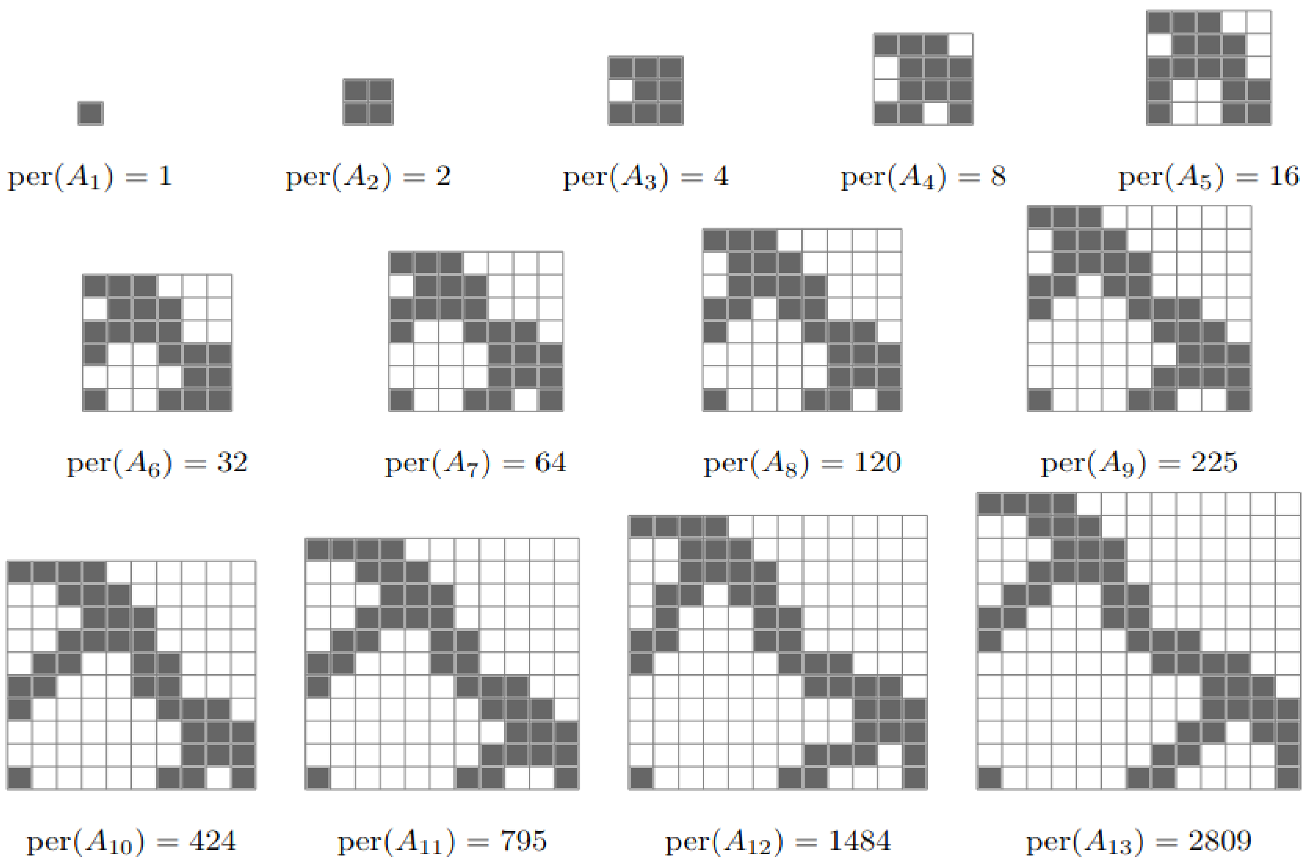
Permanent in pattern avoiding 0-1-matrices

Using the cross-entropy method, find 312-avoiding matrices with high permanent:



Permanent in pattern avoiding 0-1-matrices

Using the cross-entropy method, find 312-avoiding matrices with high permanent:



Breaks conjectured value $f_{312}(5) = 12$ by Brualdi and Cao!

Permanent in pattern avoiding 0-1-matrices

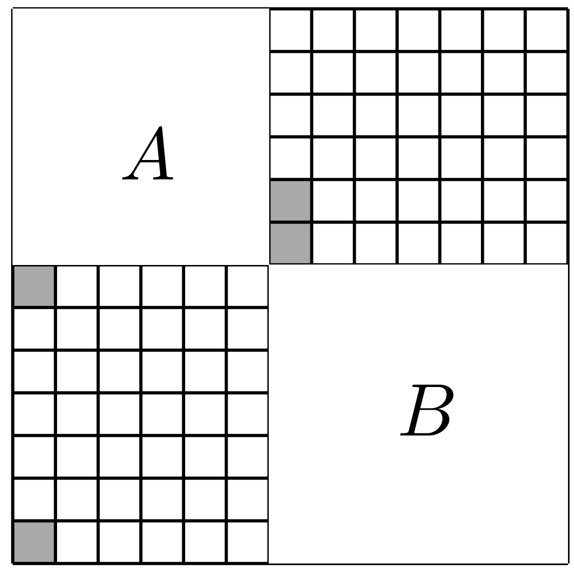
Theorem:

$$2^{0.88n} \leq f_{312}(n) \leq 24^{n/4} \approx 2^{1.15n}$$

Proof of lower bound:

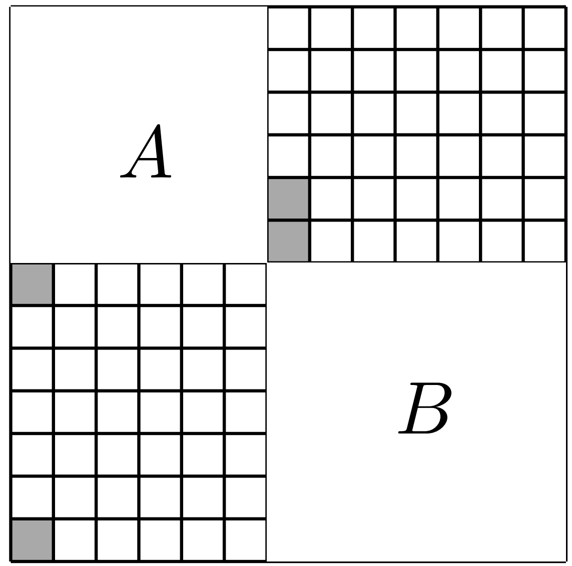
Proof of lower bound:

$A * B :=$



Proof of lower bound:

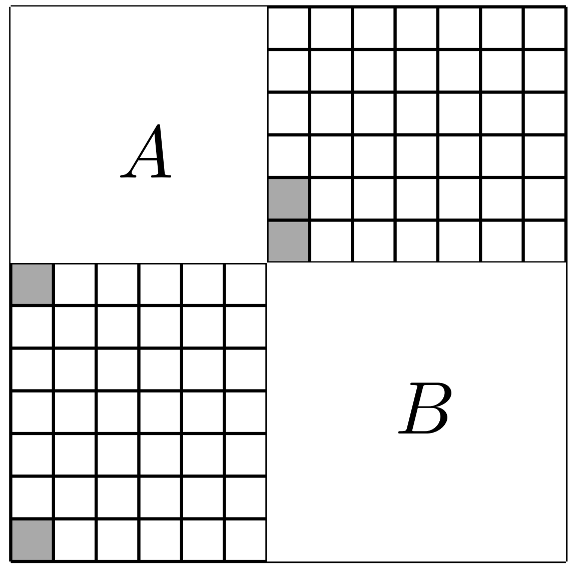
$A * B :=$



Observation:

- $\mathrm{per}(A * B) \geq \mathrm{per}(A) \cdot \mathrm{per}(B)$
- $f_{312}(n+m) \geq f_{312}(n) \cdot f_{312}(m)$

Proof of lower bound:

$A * B :=$ 

Observation:

- $\mathrm{per}(A * B) \geq \mathrm{per}(A) \cdot \mathrm{per}(B)$
- $f_{312}(n+m) \geq f_{312}(n) \cdot f_{312}(m)$

Idea: Use $\mathrm{per}(A_{13}) = 2809 > 2^{0.88 \cdot 13}$ and Fekete's Lemma.

$\square$

Proof of upper bound:

Proof of upper bound:

Theorem (Bregman & Minc, 1973):

For $A \in \{0,1\}^{n \times n}$ with row sums $r_1, ..., r_n$:

$$\mathrm{per}(A) \leq \prod_{i=1}^{n} (r_i!)^{1/r_i}$$

Proof of upper bound:

Theorem (Bregman & Minc, 1973):
For $A \in \{0,1\}^{n\times n}$ with row sums $r_1, ..., r_n$:

$$\mathrm{per}(A) \leq \prod_{i=1}^{n} (r_i!)^{1/r_i}$$

Proposition (Brualdi & Cao, 2020):
At most $4n-4$ one entries in 312-avoiding matrix $M \in \{0,1\}^{n\times n}$.

Proof of upper bound:

Theorem (Bregman & Minc, 1973):
For $A \in \{0,1\}^{n \times n}$ with row sums $r_1, \ldots, r_n$:

$$\operatorname{per}(A) \leq \prod_{i=1}^{n} (r_i!)^{1/r_i}$$

Proposition (Brualdi & Cao, 2020):
At most $4n - 4$ one entries in 312-avoiding matrix $M \in \{0,1\}^{n \times n}$.

Idea: rhs. expression maximized when all $r_i = 4$. $\square$

1-entries in pattern avoiding 0-1-matrices

1-entries in pattern avoiding 0-1-matrices

Def.: Let $f(n, \pi)$ be max. number of 1-entries in matrix $M \in \{0, 1\}^{n \times n}$ avoiding π.

1-entries in pattern avoiding 0-1-matrices

Def.: Let $f(n, \pi)$ be max. number of 1-entries in matrix $M \in \{0, 1\}^{n \times n}$ avoiding π.

Theorem (Marcus & Tardos, 2004):
For all $\pi \in S_k$ it holds $f(n, \pi) = O(n)$.

1-entries in pattern avoiding 0-1-matrices

Def.: Let $f(n,\pi)$ be max. number of 1-entries in matrix $M \in \{0,1\}^{n\times n}$ avoiding π.

Theorem (Marcus & Tardos, 2004):
For all $\pi \in S_k$ it holds $f(n,\pi) = O(n)$.

Proof:

1-entries in pattern avoiding 0-1-matrices

Def.: Let $f(n, \pi)$ be max. number of 1-entries in matrix $M \in \{0,1\}^{n \times n}$ avoiding π.

Theorem (Marcus & Tardos, 2004):
For all $\pi \in S_k$ it holds $f(n, \pi) = O(n)$.

Proof: Assume n is multiple of k^2.

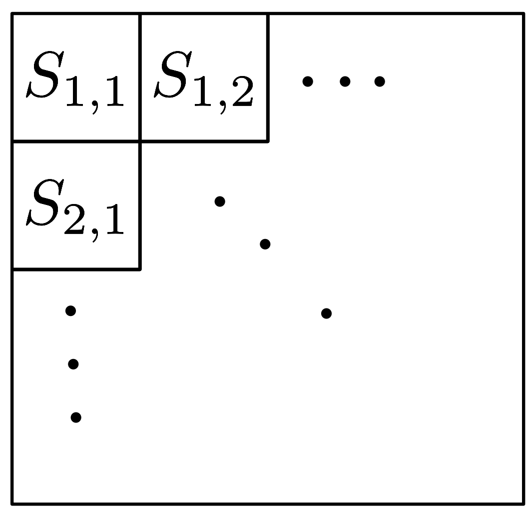
1-entries in pattern avoiding 0-1-matrices

Def.: Let $f(n, \pi)$ be max. number of 1-entries in matrix $M \in \{0,1\}^{n \times n}$ avoiding π.

Theorem (Marcus & Tardos, 2004):
For all $\pi \in S_k$ it holds $f(n, \pi) = O(n)$.

Proof: Assume n is multiple of k^2.

$M =$



$\in \{0,1\}^{n \times n}$ avoids π

Divide M into blocks $S_{i,j}$ of size $k^2 \times k^2$!

- **Def:** Block $S_{i,j}$ is *wide* (resp. *tall*),
  if at lest k columns (resp. rows) in $S_{i,j}$ contain a 1-entry.
- **Def:** Block $S_{i,j}$ is *zero*, if all entries in $S_{i,j}$ are zero.

- **Def:** Block $S_{i,j}$ is *wide* (resp. *tall*),
  if at lest k columns (resp. rows) in $S_{i,j}$ contain a 1-entry.

- **Def:** Block $S_{i,j}$ is *zero*, if all entries in $S_{i,j}$ are zero.

$$\begin{aligned} f(n,\pi) \le\ & k^4 \cdot \#\{S_{i,j} \text{ that are wide }\} \\ & + k^4 \cdot \#\{S_{i,j} \text{ that are tall}\} \\ & + (k-1)^2 \cdot \#\{S_{i,j} \text{ that are neither zero, wide, tall}\} \end{aligned}$$

- **Def:** Block $S_{i,j}$ is *wide* (resp. *tall*),
  if at lest k columns (resp. rows) in $S_{i,j}$ contain a 1-entry.

- **Def:** Block $S_{i,j}$ is *zero*, if all entries in $S_{i,j}$ are zero.

$$\begin{aligned} f(n,\pi) \leq\ & k^4 \cdot \#\{S_{i,j} \text{ that are wide } \} \\ & + k^4 \cdot \#\{S_{i,j} \text{ that are tall}\} \\ & + (k-1)^2 \cdot \#\{S_{i,j} \text{ that are neither zero, wide, tall}\} \end{aligned}$$

Idea: Bound these numbers and solve linear recursion.

Claim: $\#\{S_{i,j}$ that are not zero $\} \leq f\left(\frac{n}{k^2}, \pi\right)$

Claim: $\#\{S_{i,j}$ that are not zero $\} \leq f\left(\frac{n}{k^2}, \pi\right)$

Proof:

- Matrix $B = [b_{i,j}]$ with

$$b_{i,j} := \begin{cases} 0 & S_{i,j} \text{ is zero} \\ 1 & S_{i,j} \text{ is not zero} \end{cases}$$

  avoids π.

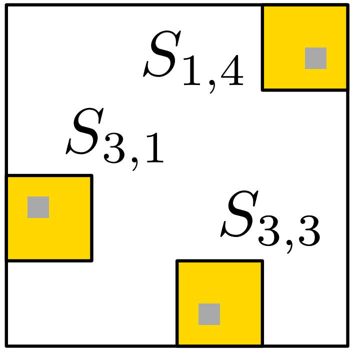
Claim: $\#\{S_{i,j}$ that are not zero $\} \leq f\left(\frac{n}{k^2}, \pi\right)$

Proof:

- Matrix $B = [b_{i,j}]$ with

$$b_{i,j} := \begin{cases} 0 & S_{i,j} \text{ is zero} \\ 1 & S_{i,j} \text{ is not zero} \end{cases}$$

avoids π.



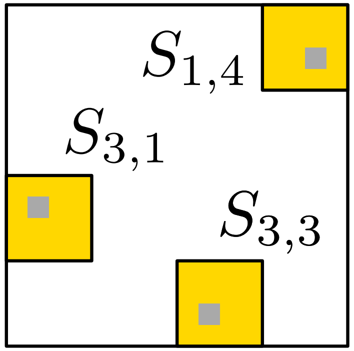
Claim: $\#\{S_{i,j}$ that are not zero $\} \leq f\left(\frac{n}{k^2}, \pi\right)$

Proof:

- Matrix $B = [b_{i,j}]$ with

$$b_{i,j} := \begin{cases} 0 & S_{i,j} \text{ is zero} \\ 1 & S_{i,j} \text{ is not zero} \end{cases}$$

  avoids π.



- $\Rightarrow$ B has at most $f\left(\frac{n}{k^2}, \pi\right)$ 1-entries. △

Claim: In each column of blocks $S_{1,j}, S_{2,j}, S_{3,j}, \cdots$, there are less than $k\binom{k^2}{k}$ wide blocks.

Claim: In each column of blocks $S_{1,j}, S_{2,j}, S_{3,j}, \cdots,$ there are less than $k\binom{k^2}{k}$ wide blocks.

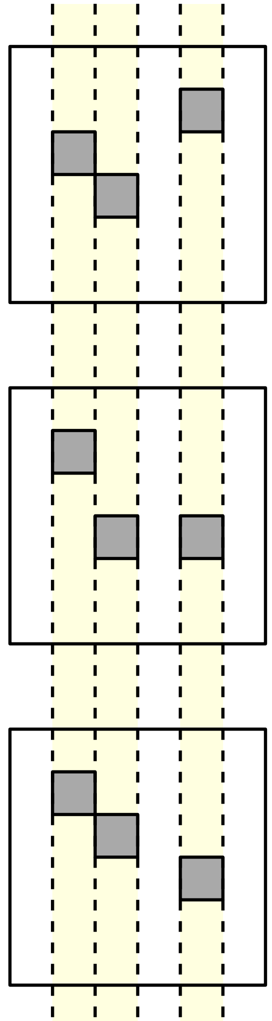
Proof:

- Suppose not, then by pigeonhole principle:
  k blocks share k columns of M
  where they all have at least one 1-entry.

Claim: In each column of blocks $S_{1,j}, S_{2,j}, S_{3,j}, \cdots$, there are less than $k\binom{k^2}{k}$ wide blocks.

Proof:

- Suppose not, then by pigeonhole principle:
  k blocks share k columns of M
  where they all have at least one 1-entry.

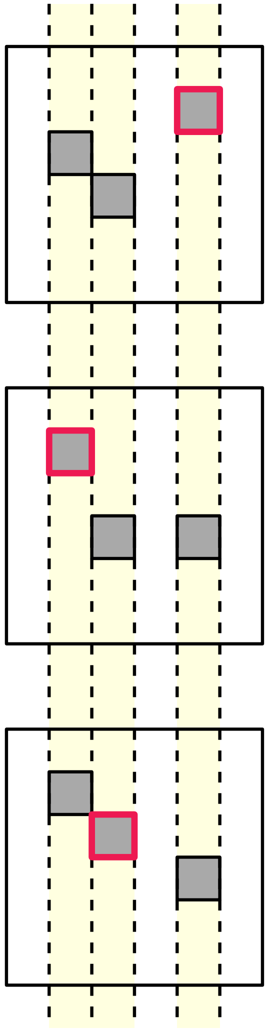


Claim: In each column of blocks $S_{1,j}, S_{2,j}, S_{3,j}, \cdots,$ there are less than $k\binom{k^2}{k}$ wide blocks.

Proof:

- Suppose not, then by pigeonhole principle:

  k blocks share k columns of M where they all have at least one 1-entry.



Claim: In each column of blocks $S_{1,j}, S_{2,j}, S_{3,j}, \cdots,$ there are less than $k\binom{k^2}{k}$ wide blocks.

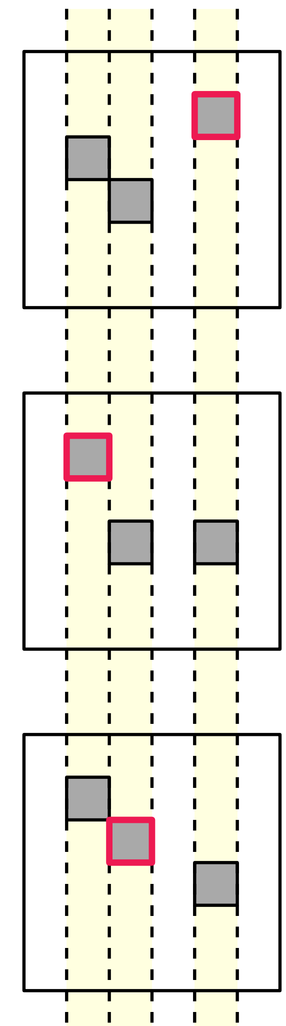
Proof:

- Suppose not, then by pigeonhole principle:

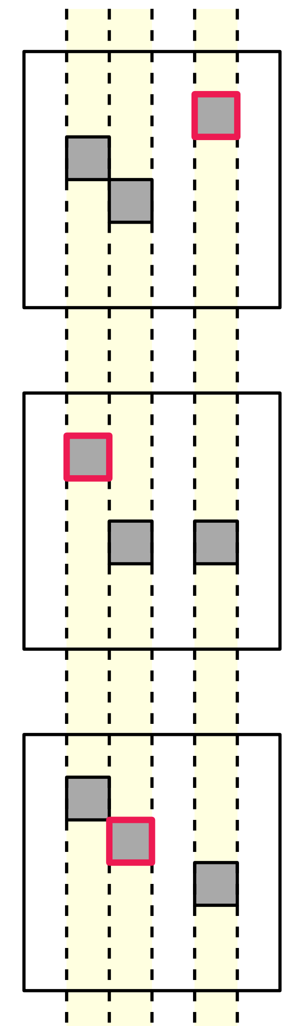
  k blocks share k columns of M
  where they all have at least one 1-entry.

- $\Rightarrow M$ contains every pattern $\pi \in S_k$ ↯

△



Claim: In each column of blocks $S_{1,j}, S_{2,j}, S_{3,j}, \cdots$, there are less than $k\binom{k^2}{k}$ wide blocks.



Proof:

- Suppose not, then by pigeonhole principle:
  k blocks share k columns of M
  where they all have at least one 1-entry.

- $\Rightarrow M$ contains every pattern $\pi \in S_k$

$\triangle$

Analogous: In each row of blocks $S_{i,1}, S_{i,2}, \cdots$ there are less than $k\binom{k^2}{k}$ tall blocks.

Conclusion:

$$
\begin{aligned}
f(n,\pi) \leq\ & k^4 \cdot \#\{S_{i,j} \text{ that are wide }\} \\
& + k^4 \cdot \#\{S_{i,j} \text{ that are tall}\} \\
& + (k-1)^2 \cdot \#\{S_{i,j} \text{ that are neither zero, wide, tall}\}
\end{aligned}
$$

Conclusion:

$$
\begin{aligned}
f(n,\pi) &\le k^4 \cdot \#\{S_{i,j} \text{ that are wide }\} \\
&+ k^4 \cdot \#\{S_{i,j} \text{ that are tall}\} \\
&+ (k-1)^2 \cdot \#\{S_{i,j} \text{ that are neither zero, wide, tall}\} \\
&\le 2k^3 \binom{k^2}{k} n + (k-1)^2 f\left(\frac{n}{k^2}, \pi\right)
\end{aligned}
$$

Conclusion:

$$\begin{aligned} f(n,\pi) &\leq k^4 \cdot \#\{S_{i,j} \text{ that are wide } \} \\ &+ k^4 \cdot \#\{S_{i,j} \text{ that are tall}\} \\ &+ (k-1)^2 \cdot \#\{S_{i,j} \text{ that are neither zero, wide, tall}\} \\ &\leq 2k^3 \binom{k^2}{k} n + (k-1)^2 f\left(\frac{n}{k^2}, \pi\right) \end{aligned}$$

Solve linear recursion...

$\square$

Don't avoid questions!