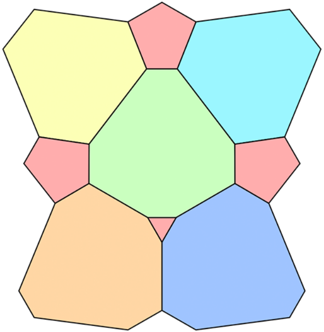
Neural Discovery in Mathematics

DOML 2025 @ RIMS

Christoph Spiegel

15th of May 2025



Results are joint work with...

Aldo Kiem
Zuse Institute Berlin
Technische U. Berlin

Konrad Mundinger
Zuse Institute Berlin
Technische U. Berlin

Sebastian Pokutta
Zuse Institute Berlin
Technische U. Berlin

Max Zimmer
Zuse Institute Berlin
Technische U. Berlin

Neural Discovery in Mathematics

What do extremal combinatorialists care about?

Extremal Graph Theory, on the most general level, investigates the extremal (maximal or minimal) value of various graph parameters over the family of graphs having a particular property. It is a lively subject with a rich history, where numerous natural questions have beautiful answers. It is a field very much driven by problems; many of the interesting ones are still wide open and stimulate an abundance of research.

Each such problem has two sides: one is the construction of an extremal structure, the other is the proof of its optimality. In this course we are putting extra emphasis on explicit constructions of extremal graphs, which do not customarily feature in standard treatments of the field. These constructions often require useful tools from algebra, geometry, or discrete Fourier analysis; the other main objective of these notes is to highlight them.

p.7 of LET'S BE EXPLICIT! lecture notes by Tibor Szabó, July 2024

Example 1: Graphs and graph sequences

Question. How many edges can a graph G of order n with $\omega(G) < r$ have?

Theorem (Mantel, 1907; Turán, 1941; Erdős–Stone, 1946)

At most $(1 - 1/r + o(1))\binom{n}{2}$, i.e., as many as the Turán graph $T(n, r)$.

Question. How large can the order n a graph G with $\max(\alpha(G), \omega(G)) < k$ be?

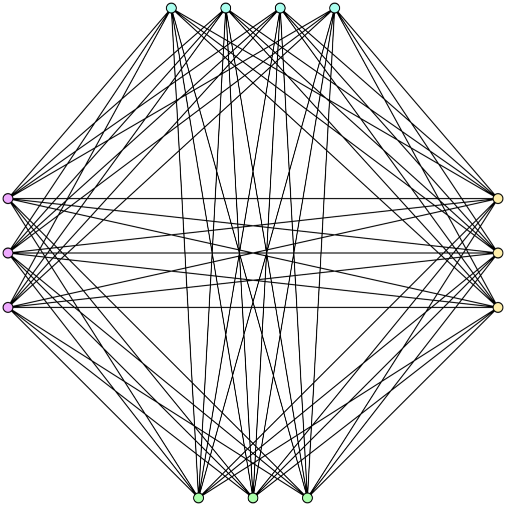
Theorem (Ramsey, 1930; many others)

We know that $R(3) = 6$, $R(4) = 18$, $43 \leq R(5) \leq 46$, and $2^{k/2} \lesssim R(k) \lesssim 3.78^k$.

A variant. How few cliques and independent sets of size r can a graph contain?

Theorem (Goodman, 1959)

Asymptotically at least 25% of all triangles need to be cliques or independent sets.



Example 1: Graphs and graph sequences

Question. How many edges can a graph G of order n with $\omega(G) < r$ have?

Theorem (Mantel, 1907; Turán, 1941; Erdős–Stone, 1946)

At most $(1 - 1/r + o(1))\binom{n}{2}$, i.e., as many as the Turán graph $T(n, r)$.

Question. How large can the order n a graph G with $\max(\alpha(G), \omega(G)) < k$ be?

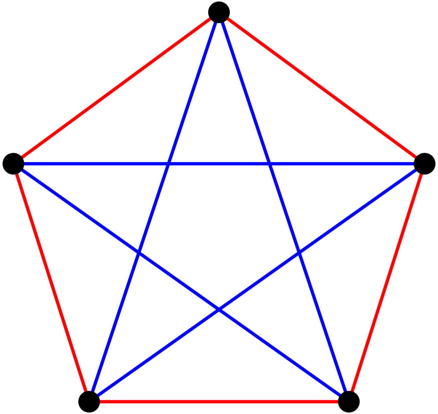
Theorem (Ramsey, 1930; many others)

We know that $R(3) = 6$, $R(4) = 18$, $43 \leq R(5) \leq 46$, and $2^{k/2} \lesssim R(k) \lesssim 3.78^k$.

A variant. How few cliques and independent sets of size r can a graph contain?

Theorem (Goodman, 1959)

Asymptotically at least 25% of all triangles need to be cliques or independent sets.



Example 1: Graphs and graph sequences

Question. How many edges can a graph G of order n with $\omega(G) < r$ have?

Theorem (Mantel, 1907; Turán, 1941; Erdős–Stone, 1946)

At most $(1 - 1/r + o(1))\binom{n}{2}$, i.e., as many as the Turán graph $T(n, r)$.

Question. How large can the order n a graph G with $\max(\alpha(G), \omega(G)) < k$ be?

Theorem (Ramsey, 1930; many others)

We know that $R(3) = 6$, $R(4) = 18$, $43 \leq R(5) \leq 46$, and $2^{k/2} \lesssim R(k) \lesssim 3.78^k$.

A variant. How few cliques and independent sets of size r can a graph contain?

Theorem (Goodman, 1959)

Asymptotically at least 25% of all triangles need to be cliques or independent sets.

Example 2: Coloring the plane

Question. What is the smallest number of colors sufficient for coloring the plane in such a way that no two points of the same color are a unit distance apart?

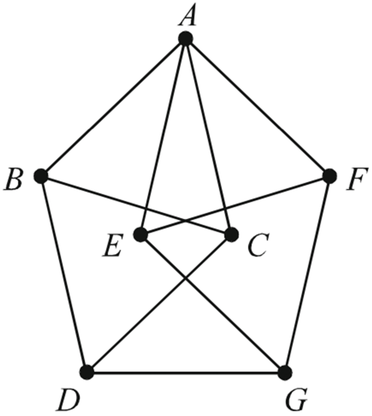
Remark. Considering the infinite graph with vertex set $\mathbb{E}^2$ and edges $\{x, y\}$ for any $\|x - y\| = 1$, we are studying the **chromatic number of the plane** $\chi(\mathbb{E}^2)$.

Theorem (Aubrey D.N.J. de Grey, 2018)

There is a unit distance graph on 20 425 *vertices with chromatic number* 5.

Upper bounds are given by explicit colorings $g : \mathbb{E}^2 \to [c] := \{1, \dots, c\}$, usually derived through tessellations using simple polytopal shapes, which give

$$5 \leq \chi(\mathbb{E}^2) \leq \dots$$


A
B
F
E
C
D
G

Example 2: Coloring the plane

Question. What is the smallest number of colors sufficient for coloring the plane in such a way that no two points of the same color are a unit distance apart?

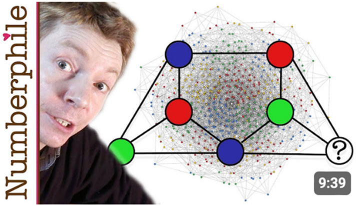
Remark. Considering the infinite graph with vertex set $\mathbb{E}^2$ and edges $\{x, y\}$ for any $\|x - y\| = 1$, we are studying the **chromatic number of the plane** $\chi(\mathbb{E}^2)$.

Theorem (Aubrey D.N.J. de Grey, 2018)

There is a unit distance graph on 20 425 *vertices with chromatic number* 5.

Upper bounds are given by explicit colorings $g : \mathbb{E}^2 \to [c] := \{1, \ldots, c\}$, usually derived through tessellations using simple polytopal shapes, which give

$$5 \leq \chi(\mathbb{E}^2) \leq \ldots$$



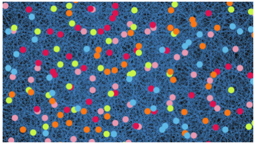
A Colorful Unsolved Problem - Numberphile

681K views • 5 years ago

Numberphile

More links & stuff in full description below ↓↓↓
Numberphile is supported by the Mathematical Science...

CC



GRAPH THEORY

Decades-Old Graph Problem Yields to Amateur Mathematician

By EVELYN LAMB | APRIL 17, 2018 | 26 |

...number of vertices? The problem, now known as the Hadwiger-Nelson problem or the problem of finding the chromatic number of the plane, has piqued the interest of many mathematicians, including...

Quantamagazine

Aubrey de Grey and Alexander Soifer, *Il Vicino*, January 18, 2020

Ronald L. Graham presents Aubrey D.N.J. de Grey the Prize: $1000, San Diego, September 22, 2018

Example 2: Coloring the plane

Question. What is the smallest number of colors sufficient for coloring the plane in such a way that no two points of the same color are a unit distance apart?

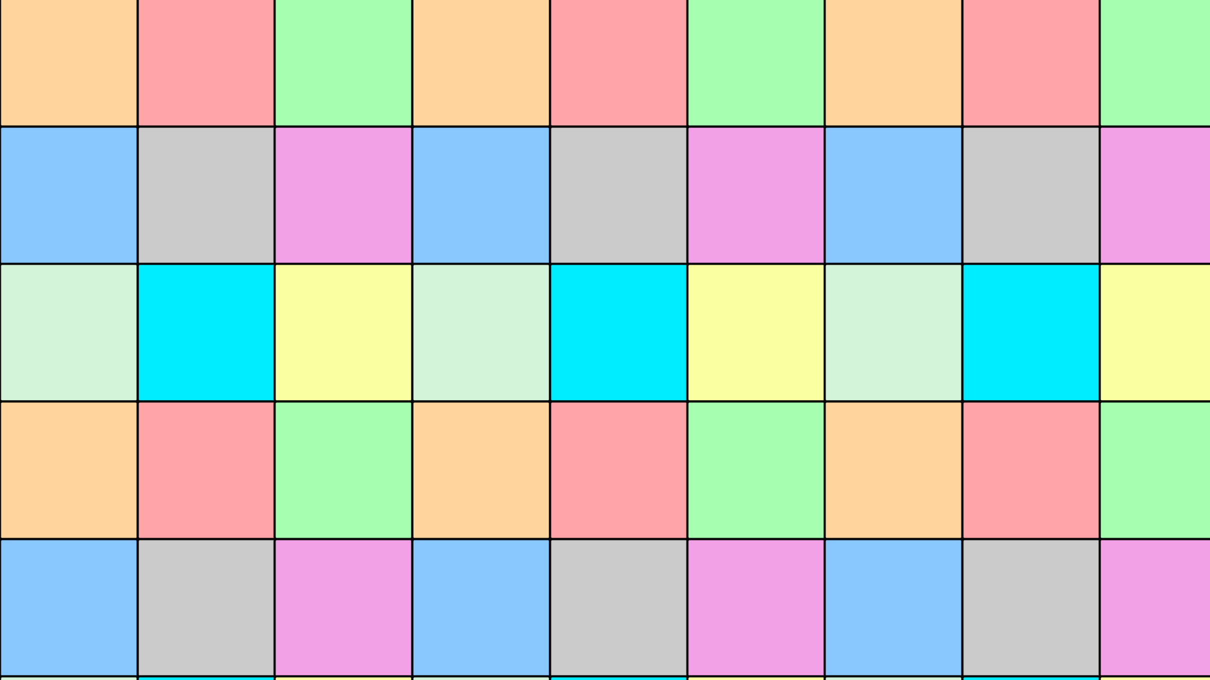
Remark. Considering the infinite graph with vertex set $\mathbb{E}^2$ and edges $\{x, y\}$ for any $\|x - y\| = 1$, we are studying the **chromatic number of the plane** $\chi(\mathbb{E}^2)$.

Theorem (Aubrey D.N.J. de Grey, 2018)

There is a unit distance graph on $20\,425$ *vertices with chromatic number* 5.

Upper bounds are given by explicit colorings $g : \mathbb{E}^2 \to [c] := \{1, \ldots, c\}$, usually derived through tessellations using simple polytopal shapes, which give

$$5 \leq \chi(\mathbb{E}^2) \leq \ldots$$



Example 2: Coloring the plane

Question. What is the smallest number of colors sufficient for coloring the plane in such a way that no two points of the same color are a unit distance apart?

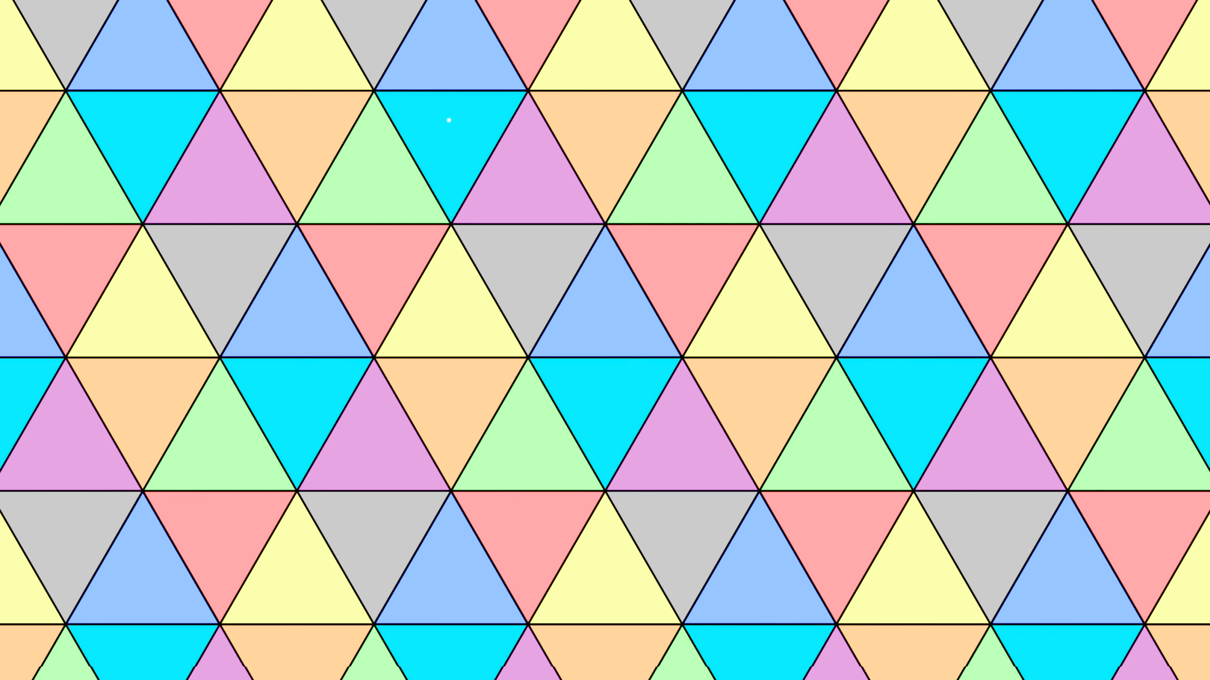
Remark. Considering the infinite graph with vertex set $\mathbb{E}^2$ and edges $\{x, y\}$ for any $\|x - y\| = 1$, we are studying the **chromatic number of the plane** $\chi(\mathbb{E}^2)$.

Theorem (Aubrey D.N.J. de Grey, 2018)

There is a unit distance graph on 20 425 *vertices with chromatic number* 5*.*

Upper bounds are given by explicit colorings $g : \mathbb{E}^2 \to [c] := \{1, \dots, c\}$, usually derived through tessellations using simple polytopal shapes, which give

$$5 \leq \chi(\mathbb{E}^2) \leq 9.$$



Example 2: Coloring the plane

Question. What is the smallest number of colors sufficient for coloring the plane in such a way that no two points of the same color are a unit distance apart?

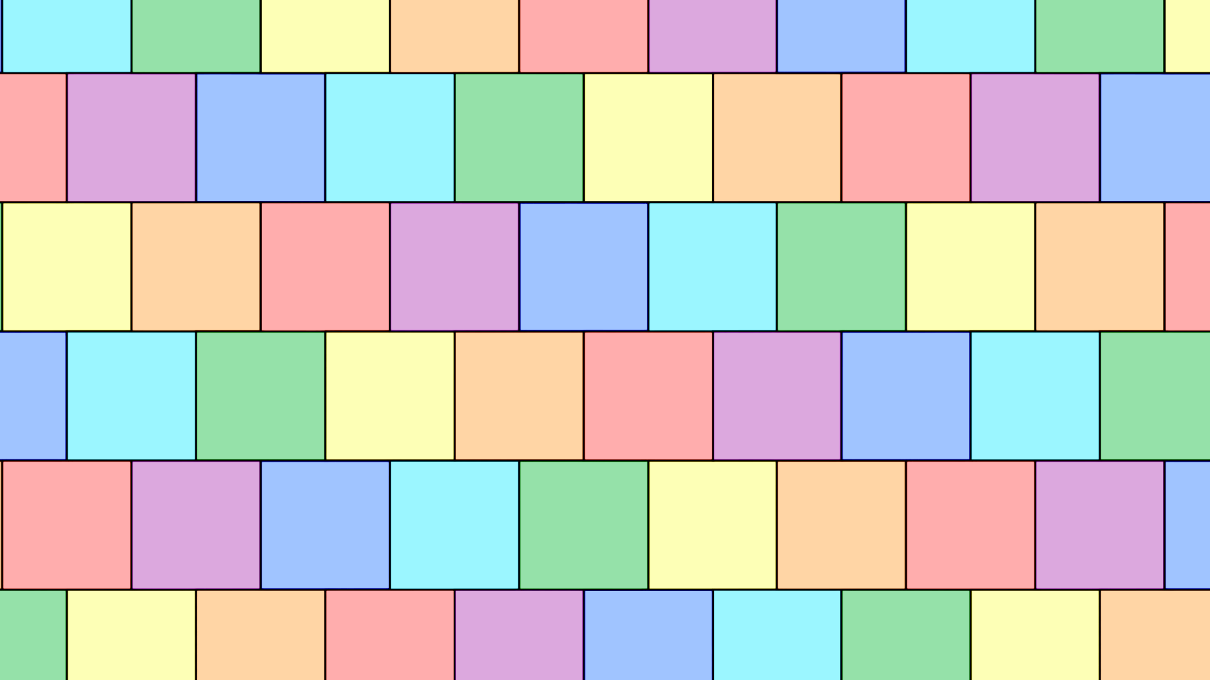
Remark. Considering the infinite graph with vertex set $\mathbb{E}^2$ and edges $\{x, y\}$ for any $\|x - y\| = 1$, we are studying the **chromatic number of the plane** $\chi(\mathbb{E}^2)$.

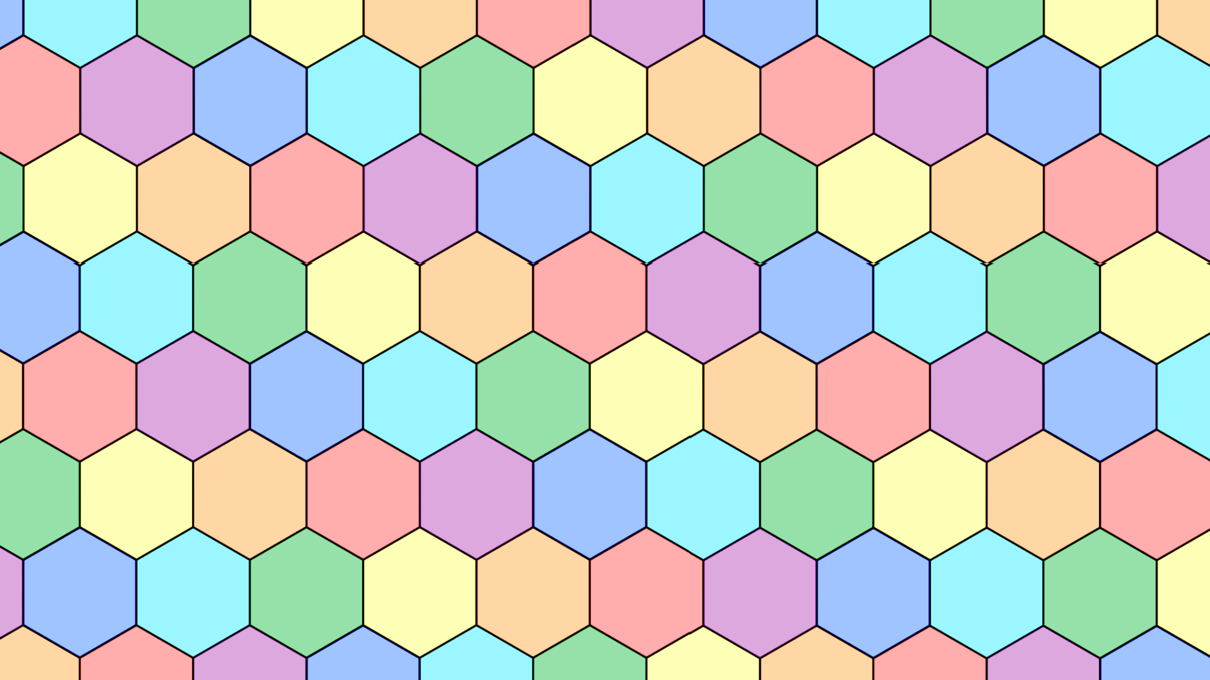
Theorem (Aubrey D.N.J. de Grey, 2018)

There is a unit distance graph on $20\,425$ *vertices with chromatic number* 5*.*

Upper bounds are given by explicit colorings $g : \mathbb{E}^2 \to [c] := \{1, \ldots, c\}$, usually derived through tessellations using simple polytopal shapes, which give

$$5 \leq \chi(\mathbb{E}^2) \leq 8.$$





Example 2: Coloring the plane

Question. What is the smallest number of colors sufficient for coloring the plane in such a way that no two points of the same color are a unit distance apart?

Remark. Considering the infinite graph with vertex set $\mathbb{E}^2$ and edges $\{x, y\}$ for any $\|x - y\| = 1$, we are studying the **chromatic number of the plane** $\chi(\mathbb{E}^2)$.

Theorem (Aubrey D.N.J. de Grey, 2018)

There is a unit distance graph on $20\,425$ *vertices with chromatic number* 5.

Upper bounds are given by explicit colorings $g : \mathbb{E}^2 \to [c] := \{1, \ldots, c\}$, usually derived through tessellations using simple polytopal shapes, which give

$$5 \leq \chi(\mathbb{E}^2) \leq 7.$$

Constructions through Implicit Representation

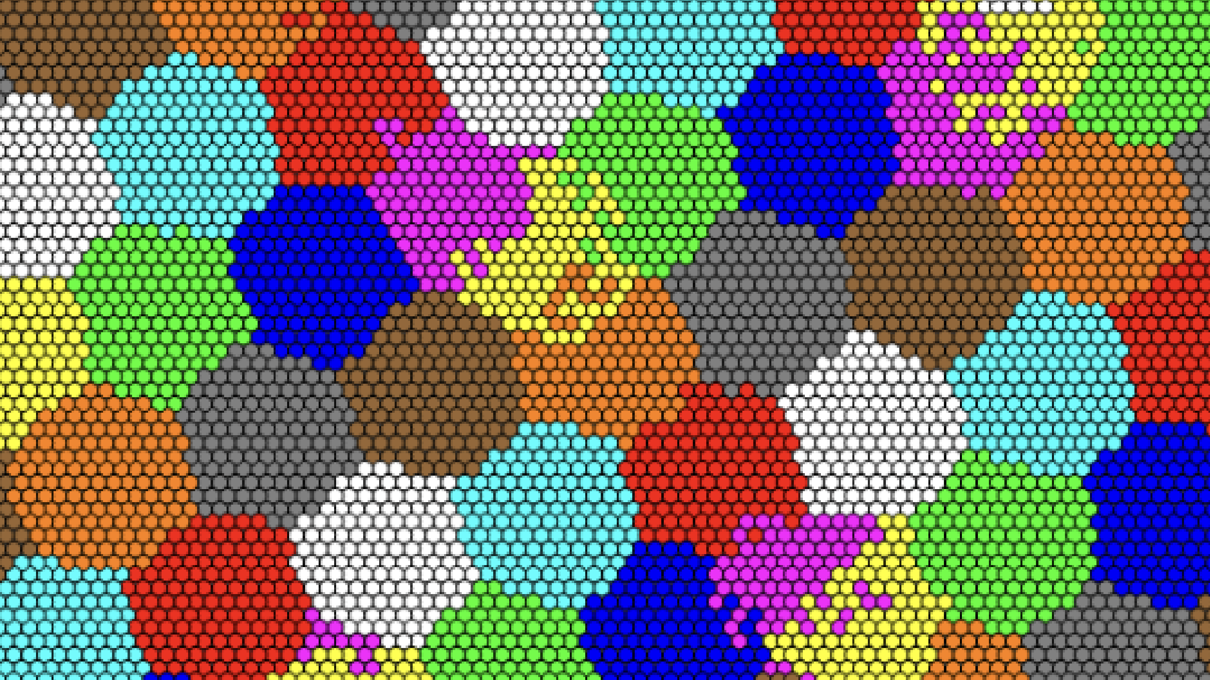
Question. Can we use computers to find colorings $g : \mathbb{E}^2 \to [c]$ so that

$$\left\{x \in \mathbb{E}^2 \mid g(x) = g(y) \text{ for any } y \in B_1(x)\right\} = \emptyset?$$

Idea. Consider a probabilistic relaxation to functions $p : \mathbb{E}^2 \to \Delta_c$ minimizing the loss

$$L_R(p) := \int_{[-R,R]^2} \int_{\partial B_1(x)} p(x)^T p(y) \, \mathrm{d}y \, \mathrm{d}x. \tag{1}$$

Challenge. Can we find a *parameterized and (easily) differentiable* family p_θ, i.e., an implicit representation, and optimize Equation (1) over θ through gradient descent?



Constructions through Implicit Representation

Question. Can we use computers to find colorings $g : \mathbb{E}^2 \to [c]$ so that

$$\left\{x \in \mathbb{E}^2 \mid g(x) = g(y) \text{ for any } y \in B_1(x)\right\} = \emptyset?$$

Idea. Consider a probabilistic relaxation to functions $p : \mathbb{E}^2 \to \Delta_c$ minimizing the loss

$$L_R(p) := \int_{[-R,R]^2} \int_{\partial B_1(x)} p(x)^T p(y) \, \mathrm{d}y \, \mathrm{d}x. \tag{1}$$

Challenge. Can we find a *parameterized and (easily) differentiable* family p_θ, i.e., an implicit representation, and optimize Equation (1) over θ through gradient descent?

Constructions through Implicit Representation

Question. Can we use computers to find colorings $g : \mathbb{E}^2 \to [c]$ so that

$$\left\{x \in \mathbb{E}^2 \mid g(x) = g(y) \text{ for any } y \in B_1(x)\right\} = \emptyset ?$$

Idea. Consider a probabilistic relaxation to functions $p : \mathbb{E}^2 \to \Delta_c$ minimizing the loss

$$L_R(p) := \int_{[-R,R]^2} \int_{\partial B_1(x)} p(x)^T p(y) \, \mathrm{d}y \, \mathrm{d}x. \qquad (1)$$

Challenge. Can we find a *parameterized and (easily) differentiable* family p_θ, i.e., an implicit representation, and optimize Equation (1) over θ through gradient descent?

How to parameterize p_θ?

A few candidates: Taylor or Fourier Series, Splines, Wavelets, SVM, RBF, ... but we chose simple feed-forward Neural Networks with sinus activations:

- Efficient gradients through backprop
- Natural spectral bias towards low frequencies (Rahaman et al., 2019)
- Implicit representation with NNs is SOTA in physics-informed learning
- Sin activation performs well for implicit repr. (Sitzmann et al., 2020)
- Mature software (`PyTorch`) and hardware (GPU) ecosystem
- Universal approximation

Theorem (Universal Approximation Theorem)

Feedforward neural networks with certain activation functions are dense (w.r.t. compact convergence) in the space of continuous functions.

How to parameterize p_θ?

A few candidates: Taylor or Fourier Series, Splines, Wavelets, SVM, RBF, ... but we chose simple feed-forward Neural Networks with sinus activations:

- Efficient gradients through backprop
- Natural spectral bias towards low frequencies (Rahaman et al., 2019)
- Implicit representation with NNs is SOTA in physics-informed learning
- Sin activation performs well for implicit repr. (Sitzmann et al., 2020)
- Mature software (`PyTorch`) and hardware (GPU) ecosystem
- Universal approximation

Theorem (Universal Approximation Theorem)

Feedforward neural networks with certain activation functions are dense (w.r.t. compact convergence) in the space of continuous functions.

How do we find the correct parameters?

Algorithm. We used *batched gradient descent* to 'train' p_θ to minimize $L_R(\theta)$. At each step we sample points $x^{(i)} \in [-R, R]^2$ and $y^{(i)} \in \partial B_1(x^{(i)})$ and use the fact that

$$\nabla_\theta L_R(\theta) \approx \hat{\nabla}_\theta L(\theta) := \sum_{i=1}^{m} \nabla_\theta\, p_\theta(x^{(i)}) \cdot p_\theta(y^{(i)})/m,$$

to adjust the parameters θ with an appropriate step size α_k through

$$\theta_{k+1} = \theta_k - \alpha_k \hat{\nabla}_\theta L(\theta).$$

Hyperparameters and implementation details.

- MLP with sinus activation functions and two hidden linear layers à 256 neurons.
- We sampled around 2^{12} pairs for each step for a total of around 2^{26} samples.
- Trained in `PyTorch` using `AdamW` and α_k linearly decaying from $\sim 10^{-3}$.

How do we find the correct parameters?

Algorithm. We used *batched gradient descent* to 'train' p_θ to minimize $L_R(\theta)$. At each step we sample points $x^{(i)} \in [-R, R]^2$ and $y^{(i)} \in \partial B_1(x^{(i)})$ and use the fact that

$$\nabla_\theta L_R(\theta) \approx \hat{\nabla}_\theta L(\theta) := \sum_{i=1}^{m} \nabla_\theta\, p_\theta(x^{(i)}) \cdot p_\theta(y^{(i)})/m,$$

to adjust the parameters θ with an appropriate step size α_k through

$$\theta_{k+1} = \theta_k - \alpha_k \hat{\nabla}_\theta L(\theta).$$

Hyperparameters and implementation details.

- MLP with sinus activation functions and two hidden linear layers à 256 neurons.
- We sampled around 2^{12} pairs for each step for a total of around 2^{26} samples.
- Trained in `PyTorch` using `AdamW` and α_k linearly decaying from $\sim 10^{-3}$.

How do we find the correct parameters?

Algorithm. We used *batched gradient descent* to 'train' p_θ to minimize $L_R(\theta)$. At each step we sample points $x^{(i)} \in [-R, R]^2$ and $y^{(i)} \in \partial B_1(x^{(i)})$ and use the fact that

$$\nabla_\theta L_R(\theta) \approx \hat{\nabla}_\theta L(\theta) := \sum_{i=1}^{m} \nabla_\theta \, p_\theta(x^{(i)}) \cdot p_\theta(y^{(i)})/m,$$

to adjust the parameters θ with an appropriate step size α_k through

$$\theta_{k+1} = \theta_k - \alpha_k \, \hat{\nabla}_\theta L(\theta).$$

Hyperparameters and implementation details.

- MLP with sinus activation functions and two hidden linear layers à 256 neurons.
- We sampled around 2^{12} pairs for each step for a total of around 2^{26} samples.
- Trained in `PyTorch` using `AdamW` and α_k linearly decaying from $\sim 10^{-3}$.

How do we find the correct parameters?

Algorithm. We used *batched gradient descent* to ‘train’ p_θ to minimize $L_R(\theta)$. At each step we sample points $x^{(i)} \in [-R, R]^2$ and $y^{(i)} \in \partial B_1(x^{(i)})$ and use the fact that

$$\nabla_\theta L_R(\theta) \approx \hat{\nabla}_\theta L(\theta) := \sum_{i=1}^{m} \nabla_\theta\, p_\theta(x^{(i)}) \cdot p_\theta(y^{(i)})/m,$$

to adjust the parameters θ with an appropriate step size α_k through

$$\theta_{k+1} = \theta_k - \alpha_k\, \hat{\nabla}_\theta L(\theta).$$

Hyperparameters and implementation details.

- MLP with sinus activation functions and two hidden linear layers à 256 neurons.
- We sampled around 2^{12} pairs for each step for a total of around 2^{26} samples.
- Trained in `PyTorch` using `AdamW` and α_k linearly decaying from $\sim 10^{-3}$.

Unfortunately this coloring was already known...

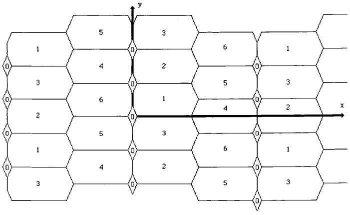
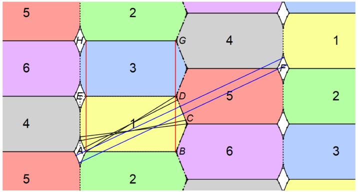
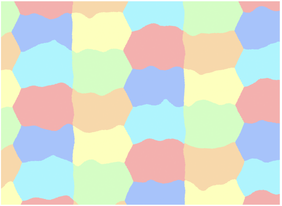


FIG. 3. A good 7-coloring of $(\mathbb{R}^2, 1)$.





Theorem (Pritikin 1995; refined by Parts 2020)

99.985% of the plane can be colored with 6 colors while avoiding unit distances.

This implies that any unit distance graph with chromatic number 7 must have order $\geq 6\,993$.

But the principle works! Can we study some variants of the original problem?

Unfortunately this coloring was already known...

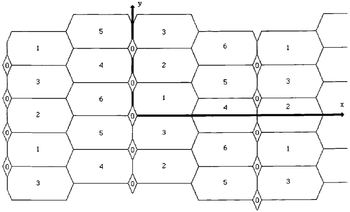
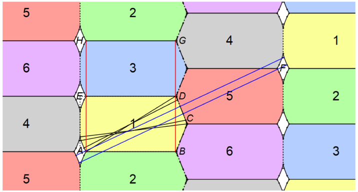
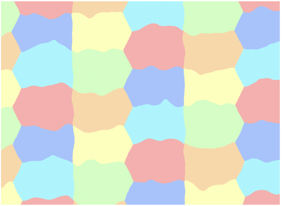


FIG. 3. A good 7-coloring of $(\mathbb{R}^2, 1)$.





Theorem (Pritikin 1995; refined by Parts 2020)

99.985% of the plane can be colored with 6 colors while avoiding unit distances.

This implies that any unit distance graph with chromatic number 7 must have order $\geq 6\,993$.

But the principle works! Can we study some variants of the original problem?

Neural Discovery in Mathematics

Variant 1: Almost succeeding...

Question. What is the smallest percentage of the plane that needs to be removed so that we can color the rest with $1, 2, \ldots, 6$ colors without monochromatic conflicts?

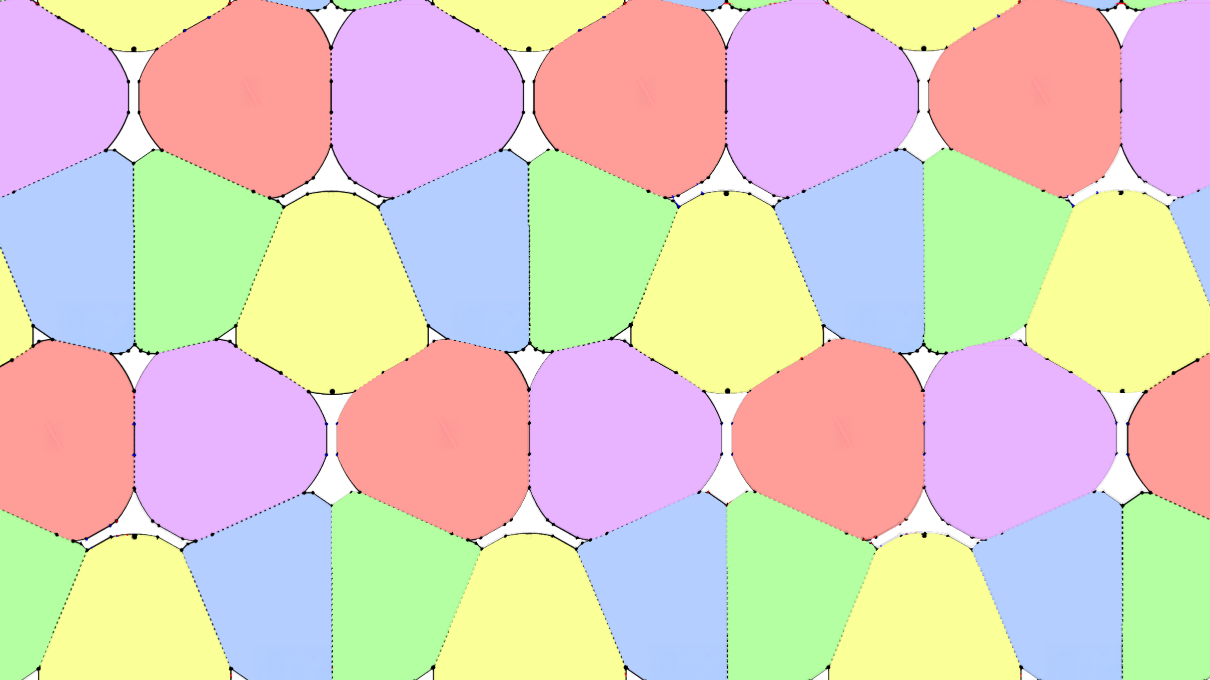
colors	**1**	**2**	**3**	**4**	**5**	**6**
best known	74.56%	54.13%	31.20%	8.25%	4.01%	0.02%
numerics	75.86%	54.14%	31.23%	8.27%	**3.56%**	0.02%

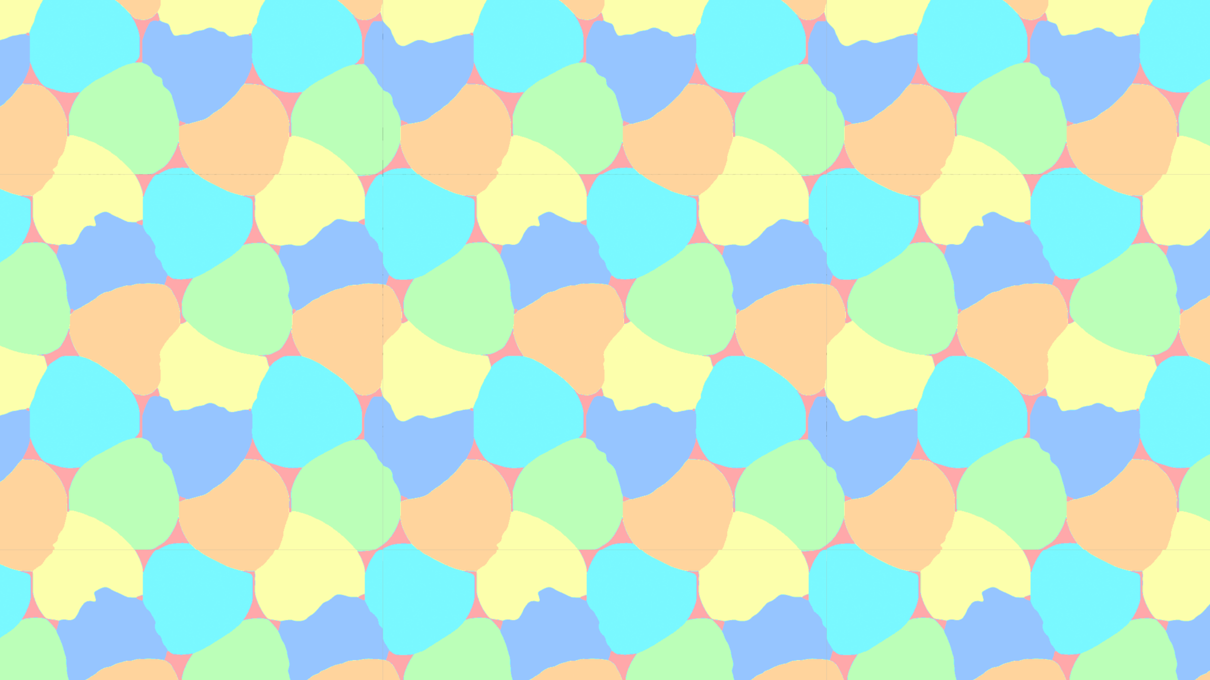
All previously best known values are due to Paarts (2020) building on work of Pritikin (1998) for 6 colors and Croft (1967) for 1, 2, 3, and 4 colors.

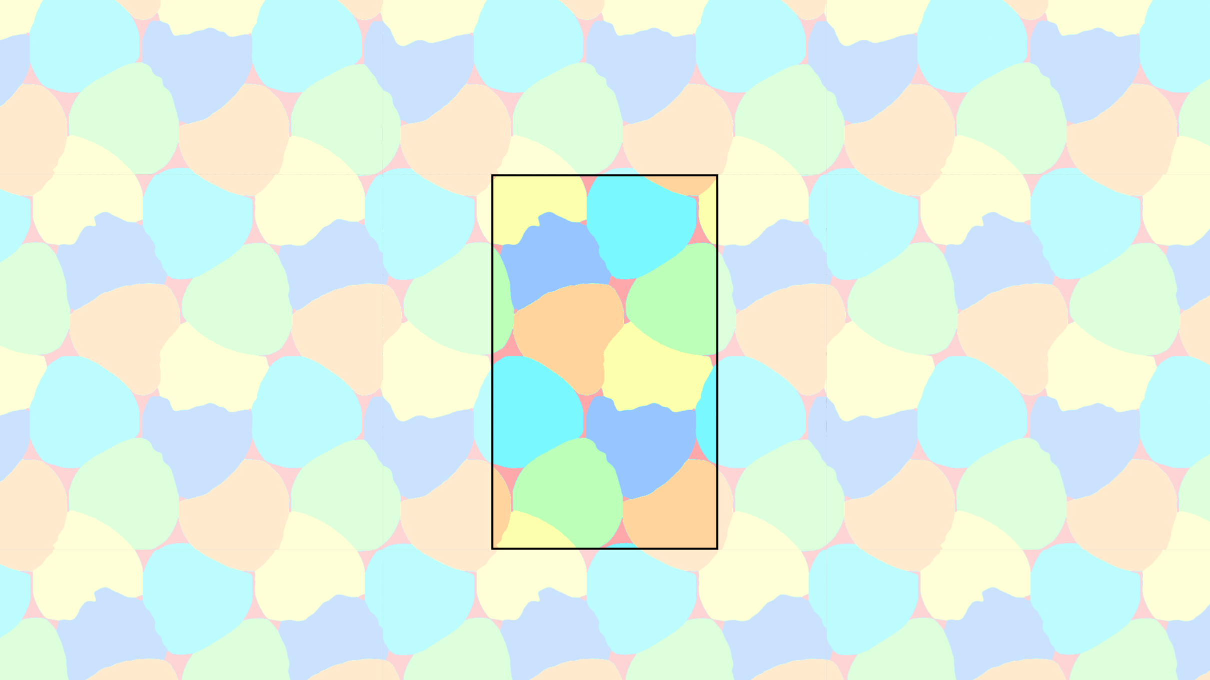
Theorem (Mundinger, Pokutta, S., Zimmer 2025+)

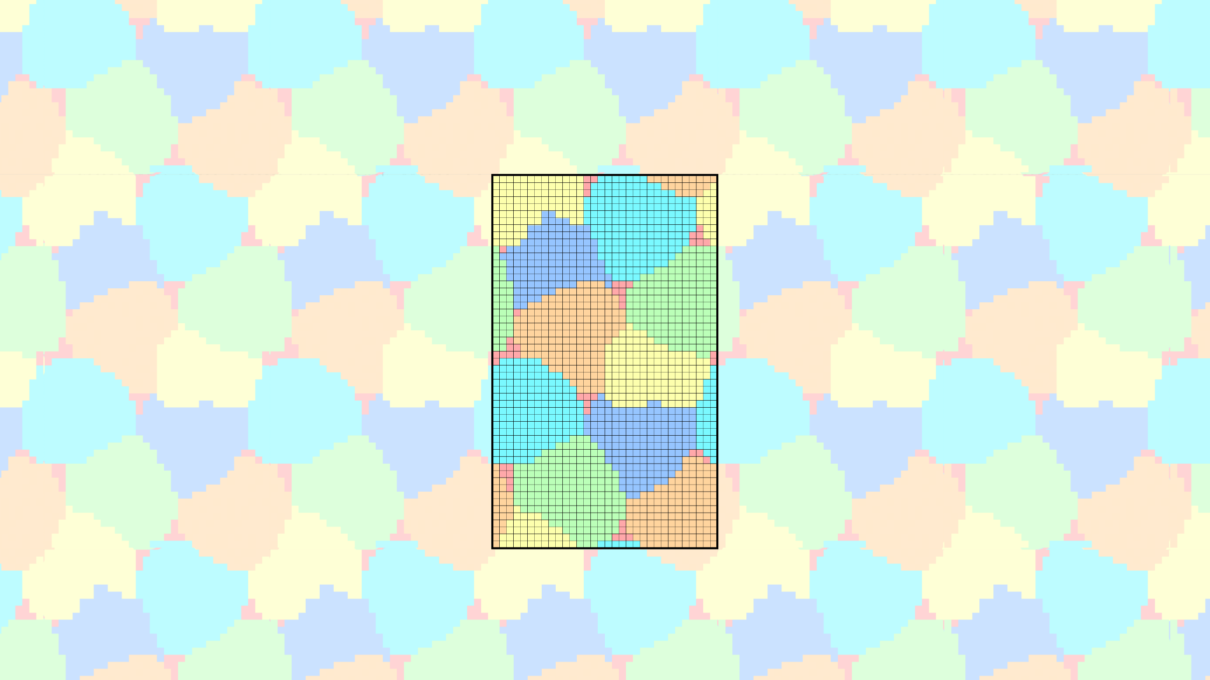
96.29% of the plane can be 5-colored with no monochromatic unit distance pairs.

Remark. We can also color $\sim 95\%$ of $\mathbb{E}^3$ using 14 colors (not yet formalized).









Variant 1: Almost succeeding...

Question. What is the smallest percentage of the plane that needs to be removed so that we can color the rest with $1, 2, \ldots, 6$ colors without monochromatic conflicts?

colors	1	2	3	4	5	6
best known	74.56%	54.13%	31.20%	8.25%	4.01%	0.02%
numerics	75.86%	54.14%	31.23%	8.27%	**3.56%**	0.02%

All previously best known values are due to Paarts (2020) building on work of Pritikin (1998) for 6 colors and Croft (1967) for 1, 2, 3, and 4 colors.

Theorem (Mundinger, Pokutta, S., Zimmer 2025+)

96.29% of the plane can be 5-colored with no monochromatic unit distance pairs.

Remark. We can also color $\sim 95\%$ of $\mathbb{E}^3$ using 14 colors (not yet formalized).

Variant 2: Going off-diagonal...

A c-coloring *realizes* $(d_1, \ldots, d_c)$ if color i does not contain distance d_i.

Problem (The continuum of six-colorings; Soifer in Nash and Rassias' *Open Problems in Mathematics*)

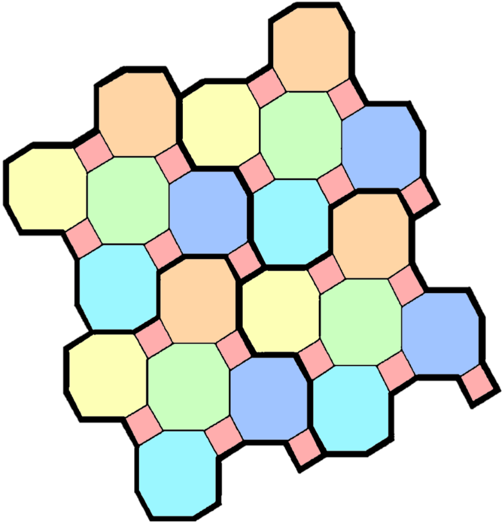
Determine the set of d for which $(1, 1, 1, 1, 1, d)$ *can be realized.*

Soifer (1991) found a coloring for $d = 1/\sqrt{5}$. Hoffman and Soifer (1993) also found one for $d = \sqrt{2} - 1$. Both of these are part of a family that covers any

$$0.414 \approx \sqrt{2} - 1 \leq d \leq 1/\sqrt{5} \approx 0.447.$$

Theorem (Mundinger, Pokutta, S., Zimmer 2024)

We extended the range of realizable types to $0.354 \leq d \leq 0.553$.



Variant 2: Going off-diagonal...

A c-coloring *realizes* $(d_1, \ldots, d_c)$ if color i does not contain distance d_i.

Problem (The continuum of six-colorings; Soifer in Nash and Rassias' *Open Problems in Mathematics*)

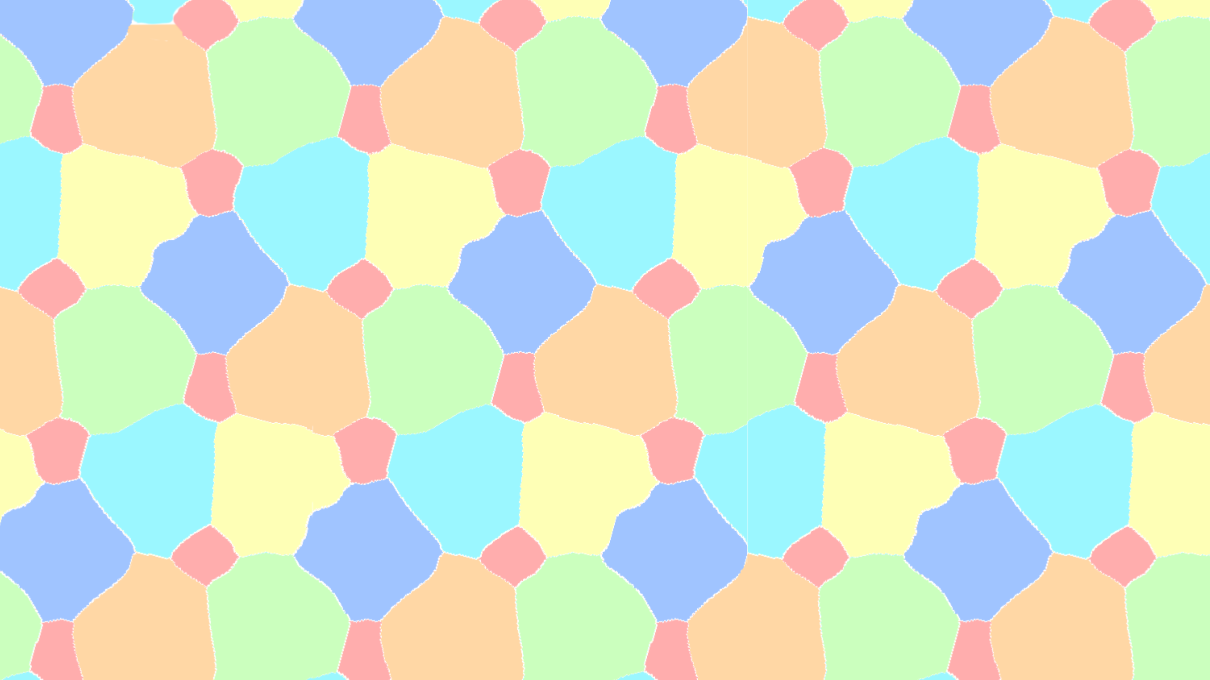
Determine the set of d for which $(1, 1, 1, 1, 1, d)$ can be realized.

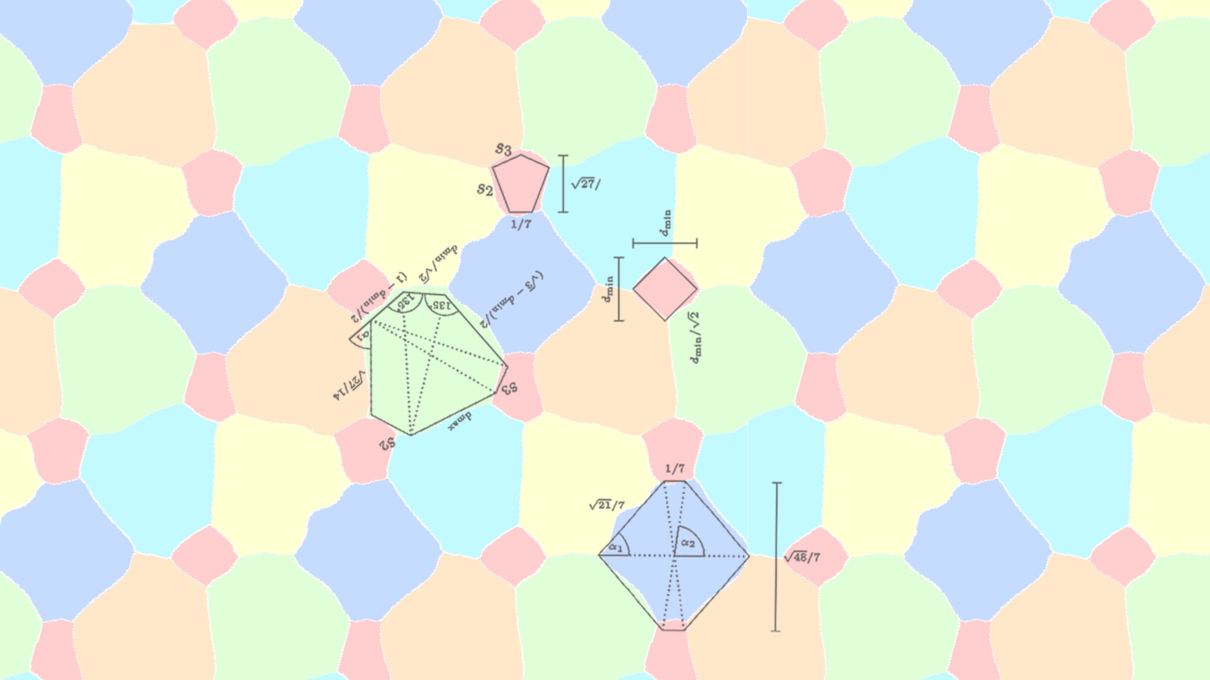
Soifer (1991) found a coloring for $d = 1/\sqrt{5}$. Hoffman and Soifer (1993) also found one for $d = \sqrt{2} - 1$. Both of these are part of a family that covers any

$$0.414 \approx \sqrt{2} - 1 \leq d \leq 1/\sqrt{5} \approx 0.447.$$

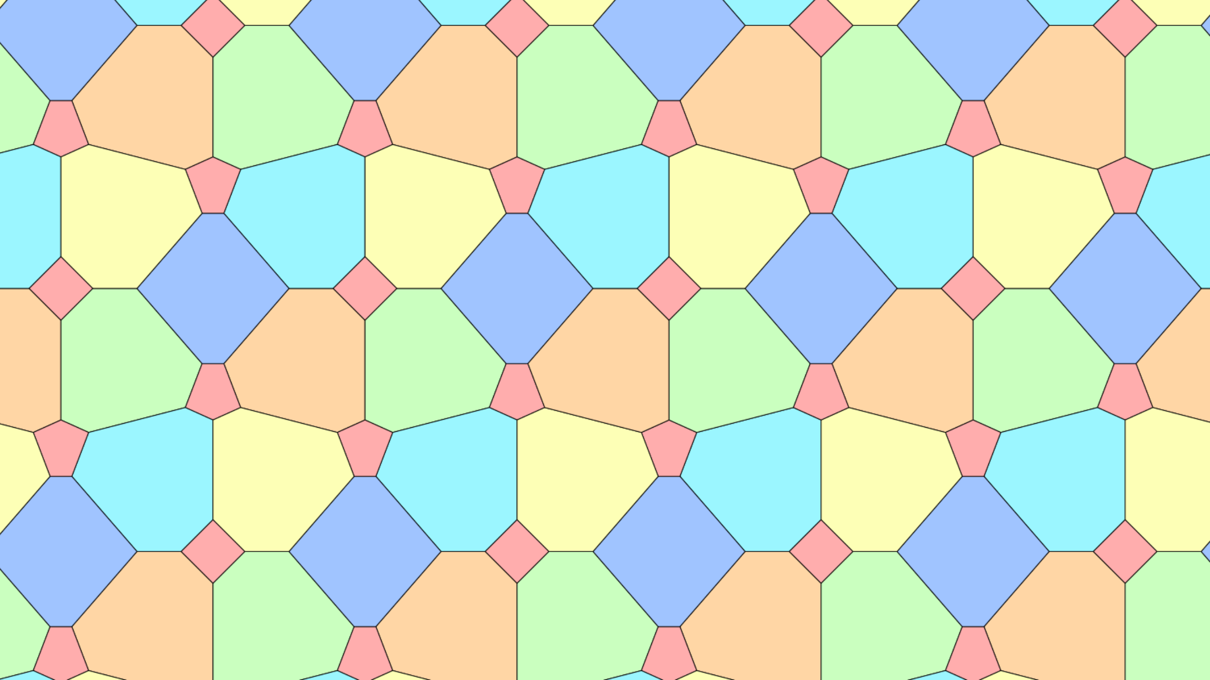
Theorem (Mundinger, Pokutta, S., Zimmer 2024)

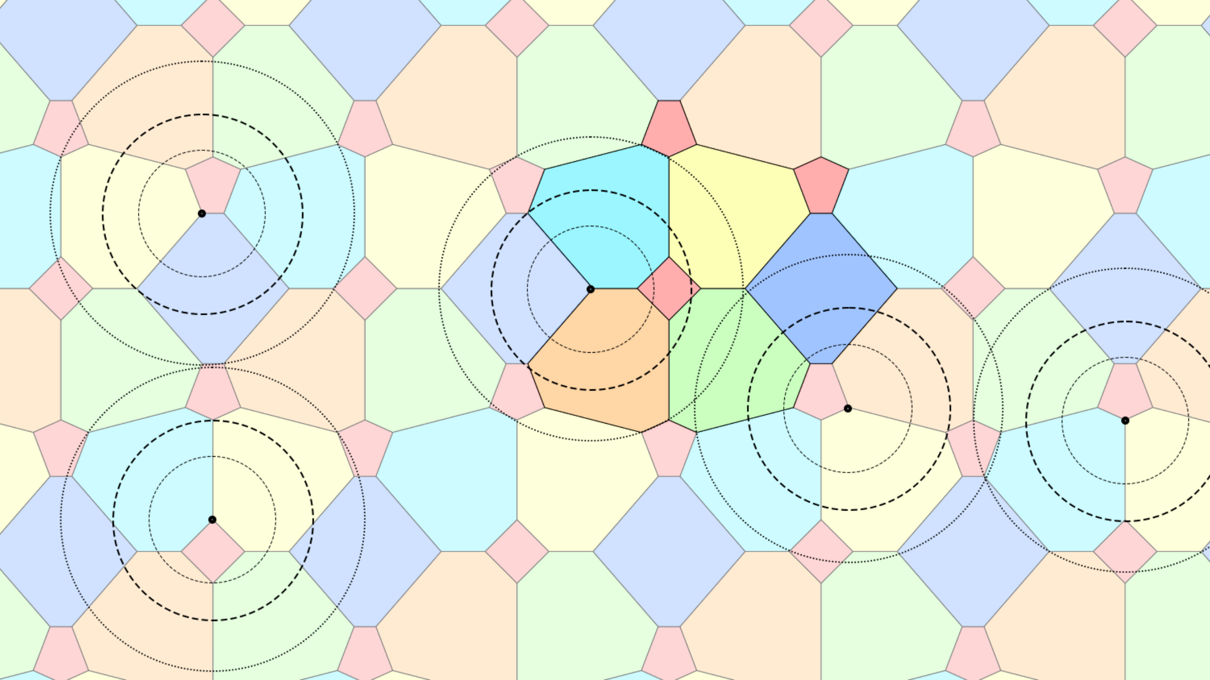
We extended the range of realizable types to $0.354 \leq d \leq 0.553$.





S3
S2
√27/
1/7
dmin
dmin
dmin/√2
dmin/√2
(1 − dmin)/2
(√3 − dmin)/2
136
135
α1
√27/14
S3
dmax
S2
1/7
√21/7
α1
α2
√48/7





Variant 2: Going off-diagonal ...

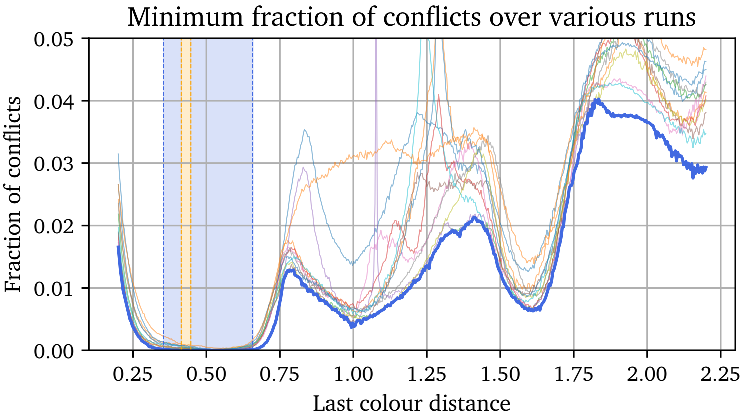
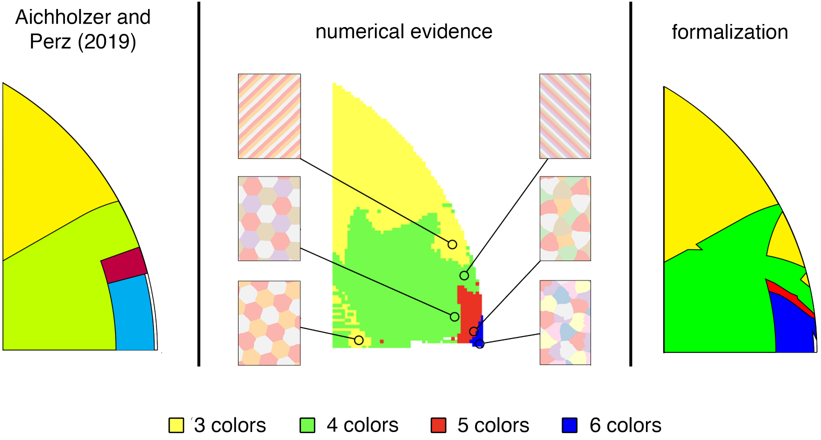


Figure: Numerical results showing the percentage of points with some conflict for a given forbidden distance in the sixth color minimized over several runs.

Variant 3: Three points are more than two...

Question. With how many colors can we color the plane while avoiding three points of the same color forming a triangles with edge lengths $0 \leq a \leq b \leq 1$?



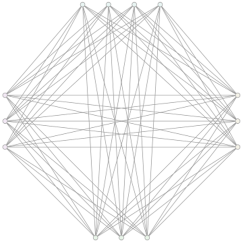
Neural Discovery in Mathematics

Graph sequences as continuous objects

Aren't graphs … discrete? Yes, but …

1) We can formulate a probabilistic relaxation through a *random graph* model.
2) *Graphons* (Lovasz and Szegedy, 2004) tell us that symmetric and measurable $\mathcal{W} : [0,1]^2 \to [0,1]$ correspond bijectively to convergent graph sequences.

You can think of graphons as adjacency matrices viewed as black-and-white images.



$$\begin{bmatrix}
 & & & & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\
 & & & & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\
 & & & & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\
 & & & & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\
1 & 1 & 1 & 1 & & & & 1 & 1 & 1 & 1 & 1 & 1 \\
1 & 1 & 1 & 1 & & & & 1 & 1 & 1 & 1 & 1 & 1 \\
1 & 1 & 1 & 1 & & & & 1 & 1 & 1 & 1 & 1 & 1 \\
1 & 1 & 1 & 1 & 1 & 1 & 1 & & & & 1 & 1 & 1 \\
1 & 1 & 1 & 1 & 1 & 1 & 1 & & & & 1 & 1 & 1 \\
1 & 1 & 1 & 1 & 1 & 1 & 1 & & & & 1 & 1 & 1 \\
1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & & & \\
1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & & & \\
1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & & &
\end{bmatrix}$$

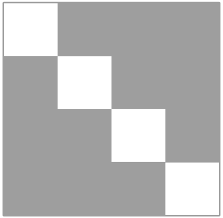


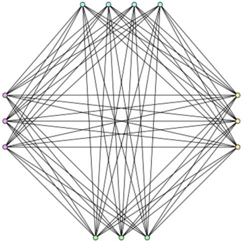
Figure: **(left)** $T(13,4)$ **(center)** adjacency matrix **(right)** graphon of $\big(T(n,4)\big)_{n\in\mathbb{N}}$

Graph sequences as continuous objects

Aren't graphs … discrete? Yes, but …

1) We can formulate a probabilistic relaxation through a *random graph* model.
2) *Graphons* (Lovasz and Szegedy, 2004) tell us that symmetric and measurable $\mathcal{W} : [0,1]^2 \to [0,1]$ correspond bijectively to convergent graph sequences.

You can think of graphons as adjacency matrices viewed as black-and-white images.



$$\begin{bmatrix}
 & & & & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\
 & & & & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\
 & & & & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\
 & & & & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\
1 & 1 & 1 & 1 & & & & 1 & 1 & 1 & 1 & 1 & 1 \\
1 & 1 & 1 & 1 & & & & 1 & 1 & 1 & 1 & 1 & 1 \\
1 & 1 & 1 & 1 & & & & 1 & 1 & 1 & 1 & 1 & 1 \\
1 & 1 & 1 & 1 & 1 & 1 & 1 & & & & 1 & 1 & 1 \\
1 & 1 & 1 & 1 & 1 & 1 & 1 & & & & 1 & 1 & 1 \\
1 & 1 & 1 & 1 & 1 & 1 & 1 & & & & 1 & 1 & 1 \\
1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & & & \\
1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & & & \\
1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & & &
\end{bmatrix}$$

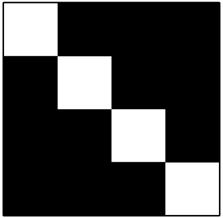


Figure: **(left)** $T(13,4)$ **(center)** adjacency matrix **(right)** graphon of $\big(T(n,4)\big)_{n\in\mathbb{N}}$

Is the same approach applicable?

Figure: Result of maximizing the number of edges while penalizing cliques of size 5 with a Lagrangian term.

Figure: Result of minimizing the number of monochromatic triangles in 3-colorings of the edges of a complete graph.

Is the same approach applicable?

Figure: Result of maximizing the number of edges while penalizing cliques of size 5 with a Lagrangian term.

Figure: Result of minimizing the number of monochromatic triangles in 3-colorings of the edges of a complete graph.

Is the same approach applicable?

Figure: Result of maximizing the number of edges while penalizing cliques of size 5 with a Lagrangian term.

Figure: Result of minimizing the number of monochromatic triangles in 3-colorings of the edges of a complete graph.

Is the same approach applicable?

Figure: Result of maximizing the number of edges while penalizing cliques of size 5 with a Lagrangian term.

Figure: Result of minimizing the number of monochromatic triangles in 3-colorings of the edges of a complete graph.

Is the same approach applicable?

Figure: Result of maximizing the number of edges while penalizing cliques of size 5 with a Lagrangian term.

Figure: Result of minimizing the number of monochromatic triangles in 3-colorings of the edges of a complete graph.

ZUSE INSTITUTE BERLIN

Thank you for your attention!

A description of the methodology was accepted at ICML 2025
and is available at arxiv.org/abs/2501.18527.

A description of the two colorings was published by Geombinatorics Quarterly
and is available at arxiv.org/abs/2404.05509.

Descriptions of the results for almost-colorings and
triangle-free colorings are in preparation.