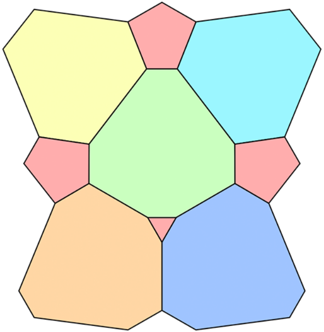
Coloring the Plane with Neural Networks

Advanced Topics in Combinatorics II @ NTU

Christoph Spiegel

23rd 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

Computational tools have a long history ...

1965 The Birch and Swinnerton-Dyer conjecture is based on **numerical evidence**

1976 Appel and Harken announce a **proof by exhaustion** of the four color theorem

1996 McCune prove that Robbins algebras are boolean using **ATP**

1998 Hales proves the Kepler conjecture by systematically solving **LPs**

2014 The Flyspeck project team announces a **formalization** of that proof

2016 Heule, Kullmann, Marek solve the Bool. Pyth. triples problem using **SAT solvers**

2017 Heule determined Schur number five in a two petabytes proof using SAT solvers

2020 The Liquid Tensor Experiment verifies a recent result of Clausen and Scholze

2024 Deepmind solves IMO problems at silver-medal level using **deep learning**

2025 Google's AlphaEvolve designs algorithms for scientific discovery

Coloring the Plane with Neural Networks

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: 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?

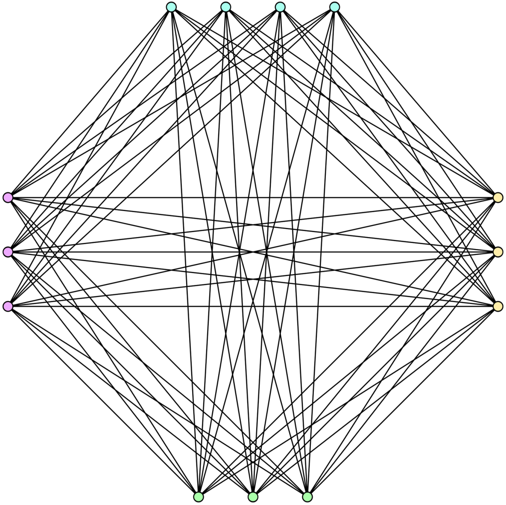
Theorem (Ramsey, 1930; many others)

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: 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?

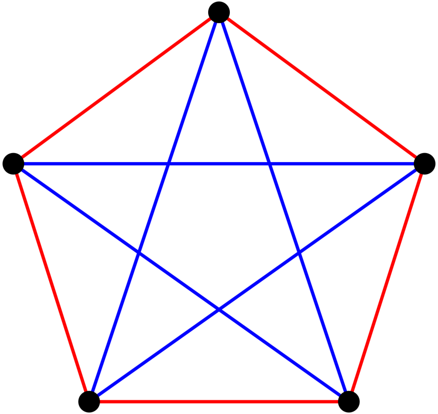
Theorem (Ramsey, 1930; many others)

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: 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.

Coloring the Plane with Neural Networks

The Hadwiger-Nelson problem

Problem (Nelson 1950, but also Hadwiger, Erdős, Gardner, Moser, Harary, Tutte, ...)

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?

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

Theorem (N.G. de Bruijn, P. Erdős 1951)

Assuming `AoC` *any graph is k-colorable iff every finite subgraph of it is k-colorable.*

This problem has a long and complicated history which has been well documented by Soifer over 14 pages in *The New Mathematical Coloring Book (2024)* ...

The Hadwiger-Nelson problem

Problem (Nelson 1950, but also Hadwiger, Erdős, Gardner, Moser, Harary, Tutte, ...)

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?

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

Theorem (N.G. de Bruijn, P. Erdős 1951)

Assuming `AoC` *any graph is k-colorable iff every finite subgraph of it is k-colorable.*

This problem has a long and complicated history which has been well documented by Soifer over 14 pages in *The New Mathematical Coloring Book (2024)* ...

The Hadwiger-Nelson problem

Problem (Nelson 1950, but also Hadwiger, Erdős, Gardner, Moser, Harary, Tutte, ...)

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?

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

Theorem (N.G. de Bruijn, P. Erdős 1951)

Assuming `AoC` *any graph is k-colorable iff every finite subgraph of it is k-colorable.*

This problem has a long and complicated history which has been well documented by Soifer over 14 pages in *The New Mathematical Coloring Book (2024)* ...

The Hadwiger-Nelson problem

Problem (Nelson 1950, but also Hadwiger, Erdős, Gardner, Moser, Harary, Tutte, ...)

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?

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

Theorem (N.G. de Bruijn, P. Erdős 1951)

Assuming `AoC` *any graph is k-colorable iff every finite subgraph of it is k-colorable.*

This problem has a long and complicated history which has been well documented by Soifer over 14 pages in *The New Mathematical Coloring Book (2024)* ...

The history of the problem

Table 3.1 Who created the chromatic number of the plane problem?

Publication	Year	Author(s)	Problem creator(s) or source named
[Gar2]	1960	Gardner	"**Leo Moser** ...writes..."
[Had4]	1961	Hadwiger (after Klee)	**Nelson**
[E61.22]	1961	Erdős	"I cannot trace the origin of this problem"
[Cro]	1967	Croft	"A long[18]-standing open problem of **Erdős**"
[Woo1]	1973	Woodall	**Gardner**
[Sim]	1976	Simmons	**Erdős**, **Harary**, and **Tutte**
[E80.38] [E81.23] [E81.26]	1980–1981	Erdős	**Hadwiger** and **Nelson**
[CFG]	1991	Croft, Falconer, and Guy	"Apparently due to **E. Nelson**"
[KW]	1991	Klee and Wagon	"Posed in 1960–61 by **M. Gardner** and **Hadwiger**"

p. 24 of The New Mathematical Coloring Book by Alexander Soifer, 2024

The history of the problem

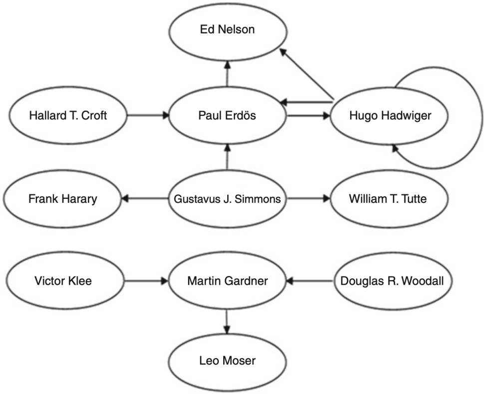


Diagram 3.1 Who created the chromatic number of the plane problem?

p. 24 of The New Mathematical Coloring Book by Alexander Soifer, 2024

The history of the problem

The results of my historical research are summarized in Diagram 3.2, where arrows show passing of the problem from one mathematician to another. In the end, Paul Erdős shares the problem with the world in numerous talks and articles.

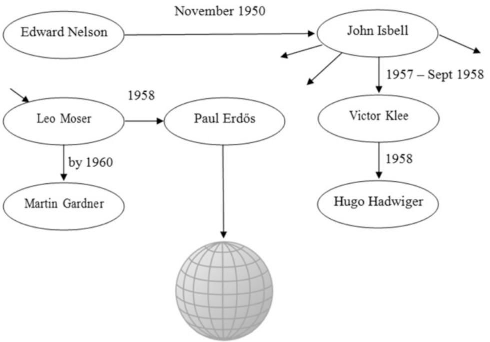


Diagram 3.2 Passing the baton of the chromatic number of the plane problem

p. 32 of The New Mathematical Coloring Book by Alexander Soifer, 2024

Bounds on the problem

Lower bounds are given by finding unit distance graphs of large chromatic number.

Definition

A graph $G = (V, E)$ is a *unit distance graph* if there exists an embedding $f : V \to \mathbb{E}^2$ of its vertices in the plane s.t. $\|f(u) - f(v)\| = 1$ if and only if $\{u, v\} \in E$.

A triangle gives a lower bound of 3 and the Moser spindle a lower bound of 4 (1961).

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 tesselations using simple polytopal shapes, which give

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

Bounds on the problem

Lower bounds are given by finding unit distance graphs of large chromatic number.

Definition

A graph $G = (V, E)$ is a *unit distance graph* if there exists an embedding $f : V \to \mathbb{E}^2$ of its vertices in the plane s.t. $\|f(u) - f(v)\| = 1$ if and only $\{u, v\} \in E$.

A triangle gives a lower bound of 3 and the Moser spindle a lower bound of 4 (1961).

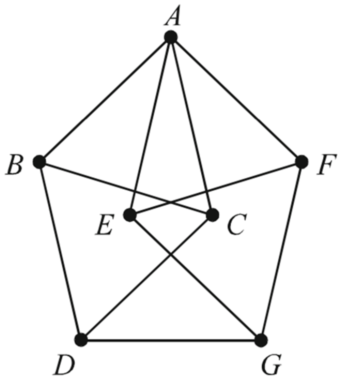
Theorem (Aubrey D.N.J. de Grey, 2018)

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 tesselations using simple polytopal shapes, which give

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

Bounds on the problem



Bounds on the problem

Lower bounds are given by finding unit distance graphs of large chromatic number.

Definition

A graph $G = (V, E)$ is a *unit distance graph* if there exists an embedding $f : V \to \mathbb{E}^2$ of its vertices in the plane s.t. $\|f(u) - f(v)\| = 1$ if and only if $\{u, v\} \in E$.

A triangle gives a lower bound of 3 and the Moser spindle a lower bound of 4 (1961).

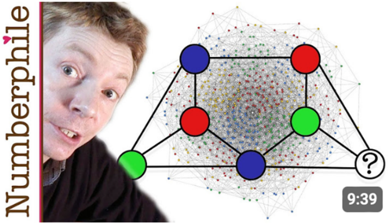
Theorem (Aubrey D.N.J. de Grey, 2018)

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 tesselations using simple polytopal shapes, which give

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

Bounds on the problem



A Colorful Unsolved Problem - Numberphile

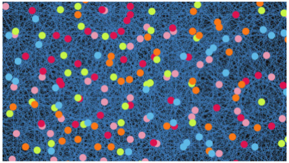
681K views • 5 years ago

Numberphile

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

CC

Bounds on the problem



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

Bounds on the problem

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

Bounds on the problem

Lower bounds are given by finding unit distance graphs of large chromatic number.

Definition

A graph $G = (V, E)$ is a *unit distance graph* if there exists an embedding $f : V \to \mathbb{E}^2$ of its vertices in the plane s.t. $\|f(u) - f(v)\| = 1$ if and only if $\{u, v\} \in E$.

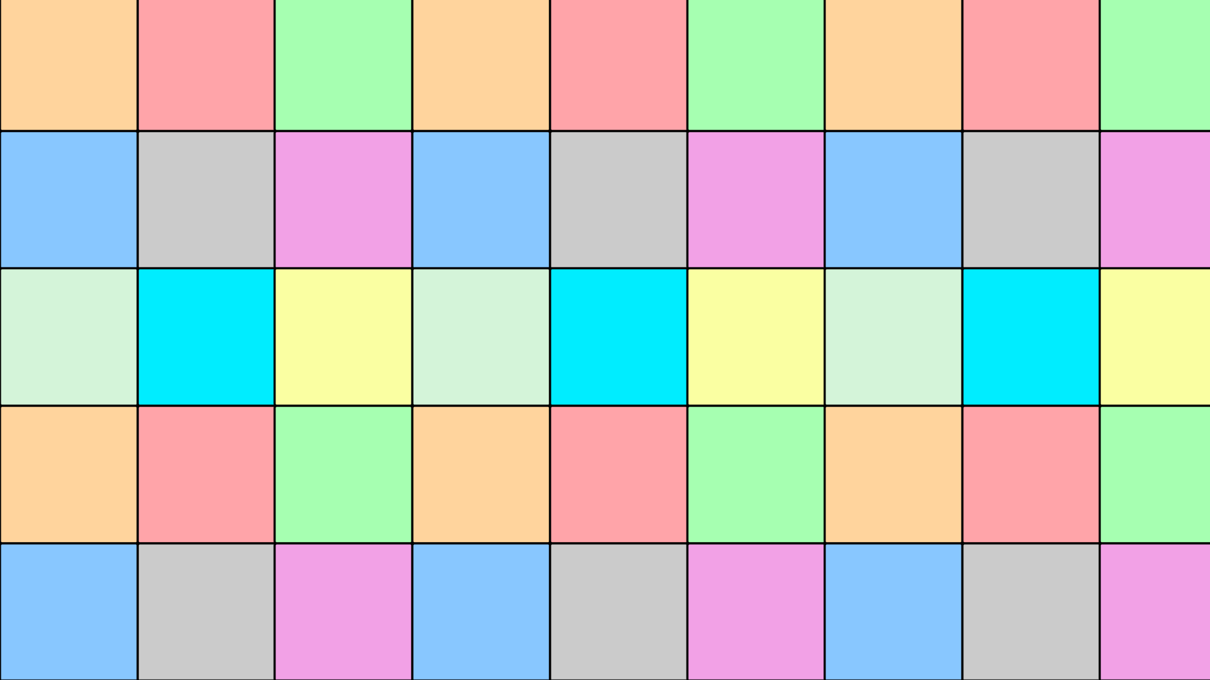
A triangle gives a lower bound of 3 and the Moser spindle a lower bound of 4 (1961).

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 tesselations using simple polytopal shapes, which give

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



Bounds on the problem

Lower bounds are given by finding unit distance graphs of large chromatic number.

Definition

A graph $G = (V, E)$ is a *unit distance graph* if there exists an embedding $f : V \to \mathbb{E}^2$ of its vertices in the plane s.t. $\|f(u) - f(v)\| = 1$ if and only if $\{u, v\} \in E$.

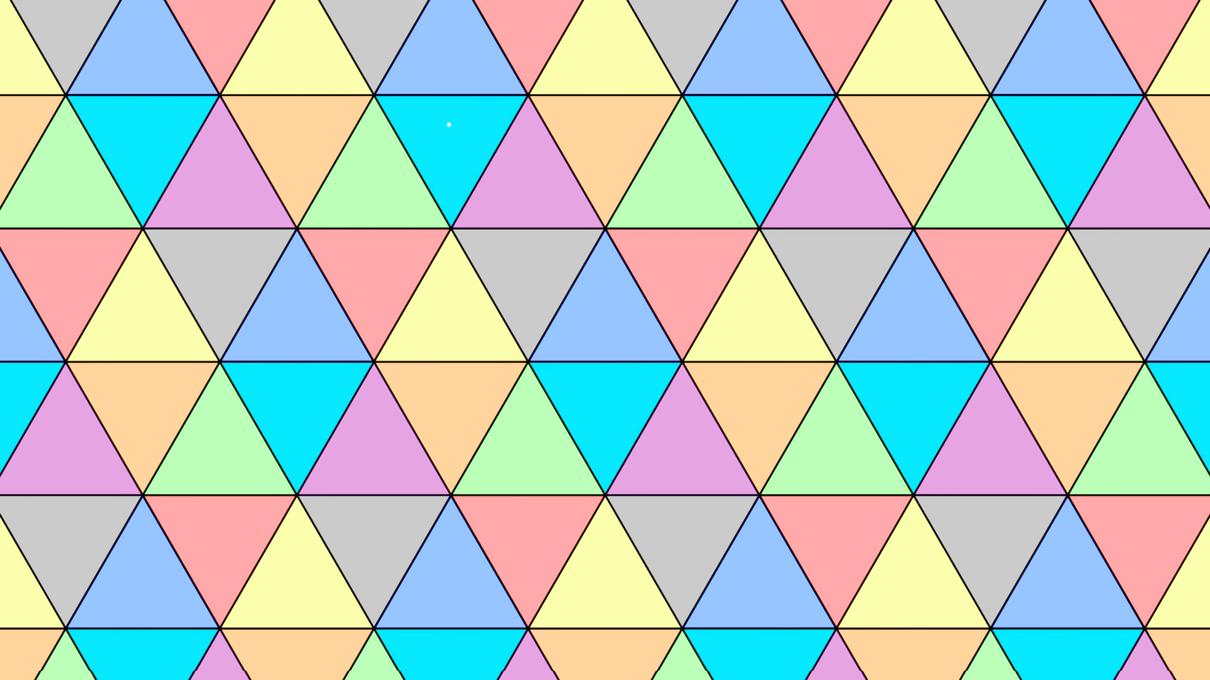
A triangle gives a lower bound of 3 and the Moser spindle a lower bound of 4 (1961).

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 tesselations using simple polytopal shapes, which give

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



Bounds on the problem

Lower bounds are given by finding unit distance graphs of large chromatic number.

Definition

A graph $G = (V, E)$ is a *unit distance graph* if there exists an embedding $f : V \to \mathbb{E}^2$ of its vertices in the plane s.t. $\|f(u) - f(v)\| = 1$ if and only if $\{u, v\} \in E$.

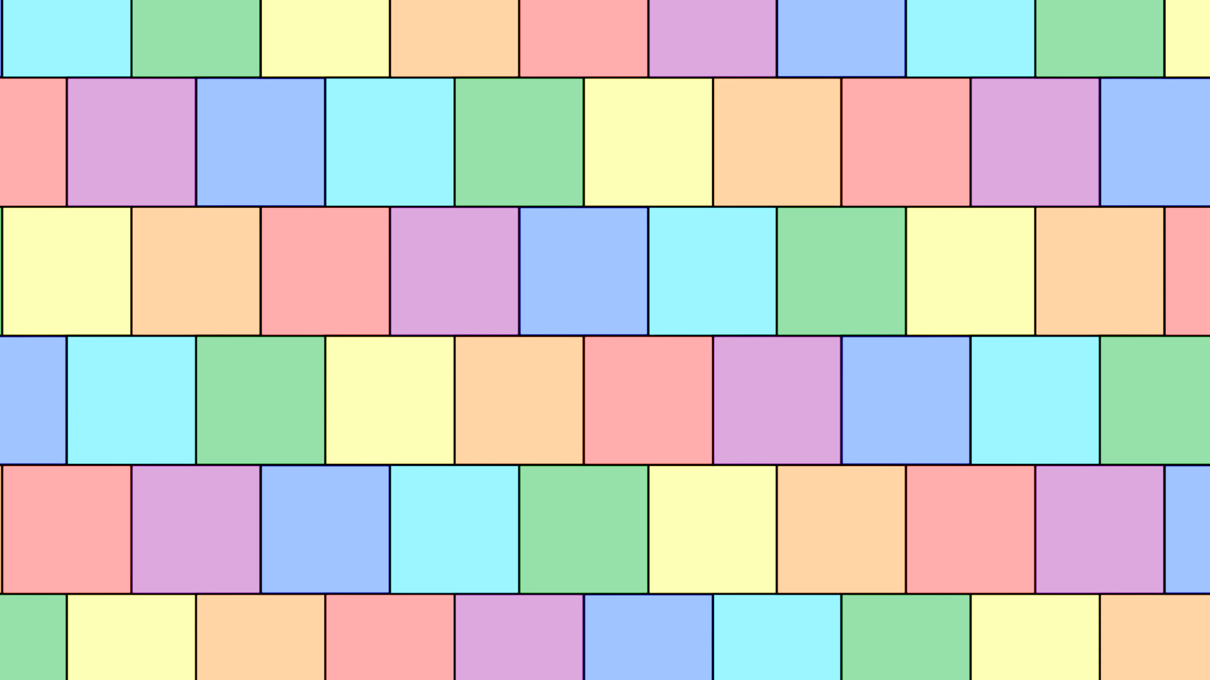
A triangle gives a lower bound of 3 and the Moser spindle a lower bound of 4 (1961).

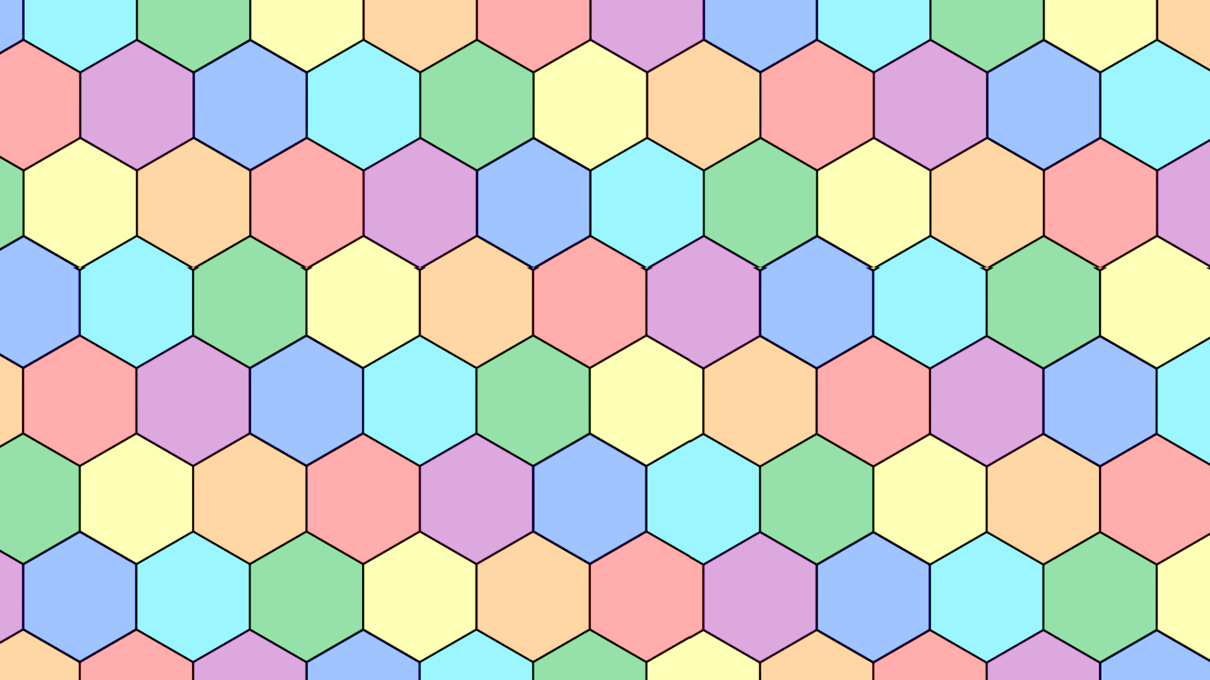
Theorem (Aubrey D.N.J. de Grey, 2018)

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 tesselations using simple polytopal shapes, which give

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





Bounds on the problem

Lower bounds are given by finding unit distance graphs of large chromatic number.

Definition

A graph $G = (V, E)$ is a *unit distance graph* if there exists an embedding $f : V \to \mathbb{E}^2$ of its vertices in the plane s.t. $\|f(u) - f(v)\| = 1$ if and only if $\{u, v\} \in E$.

A triangle gives a lower bound of 3 and the Moser spindle a lower bound of 4 (1961).

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 tesselations using simple polytopal shapes, which give

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

Coloring the Plane with Neural Networks

Constructions through Implicit Representation

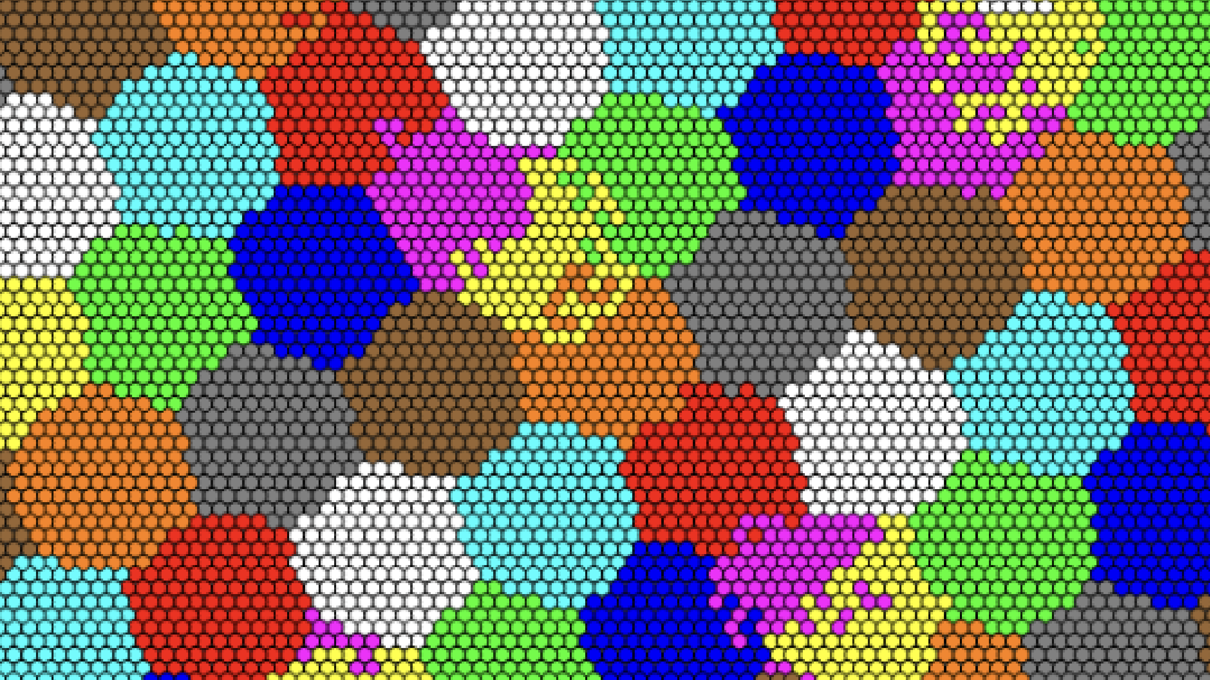
Question. Can we use computers to find colorings $g : \mathbb{E}^2 \to [c]$ so that

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

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?

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?

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:

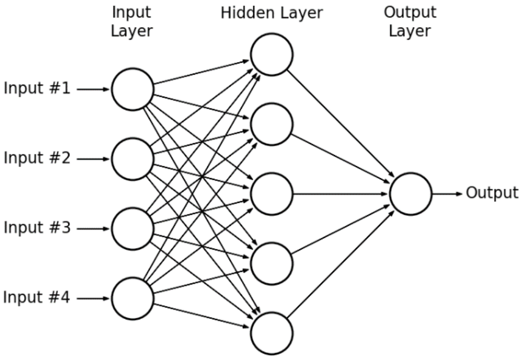


Figure: Feedforward neural network or multilayer perceptron architecture.

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:

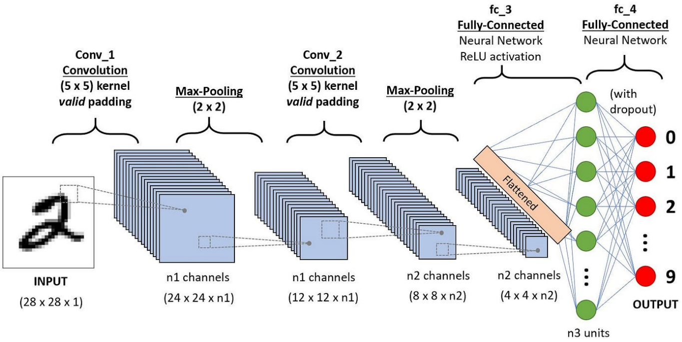


Figure: Convolutional neural network architecture.

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:

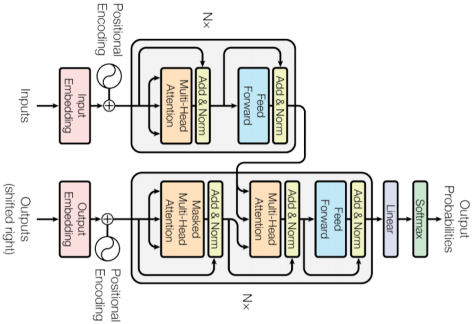


Figure: Transformer neural network architecture.

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:

input $\longrightarrow$

Blackbox

heavily parameterized
easily differentiable
universal approximator

$\longrightarrow$ output

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...

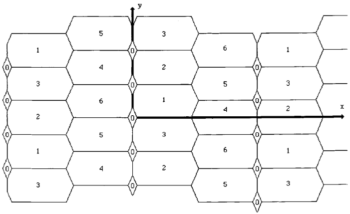
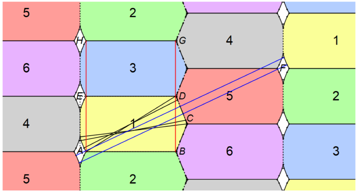
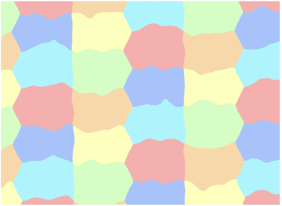


FIG. 3. A good 7-coloring of $(\mathbb{R}^2, 1)$.





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...

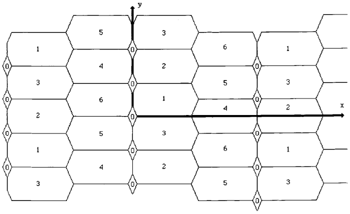
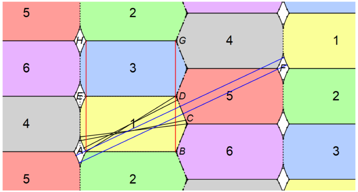
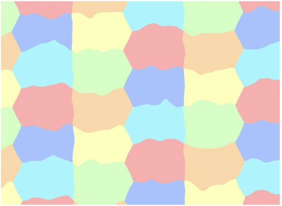


FIG. 3. A good 7-coloring of $(\mathbb{R}^2, 1)$.





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?

Coloring the Plane with Neural Networks

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?

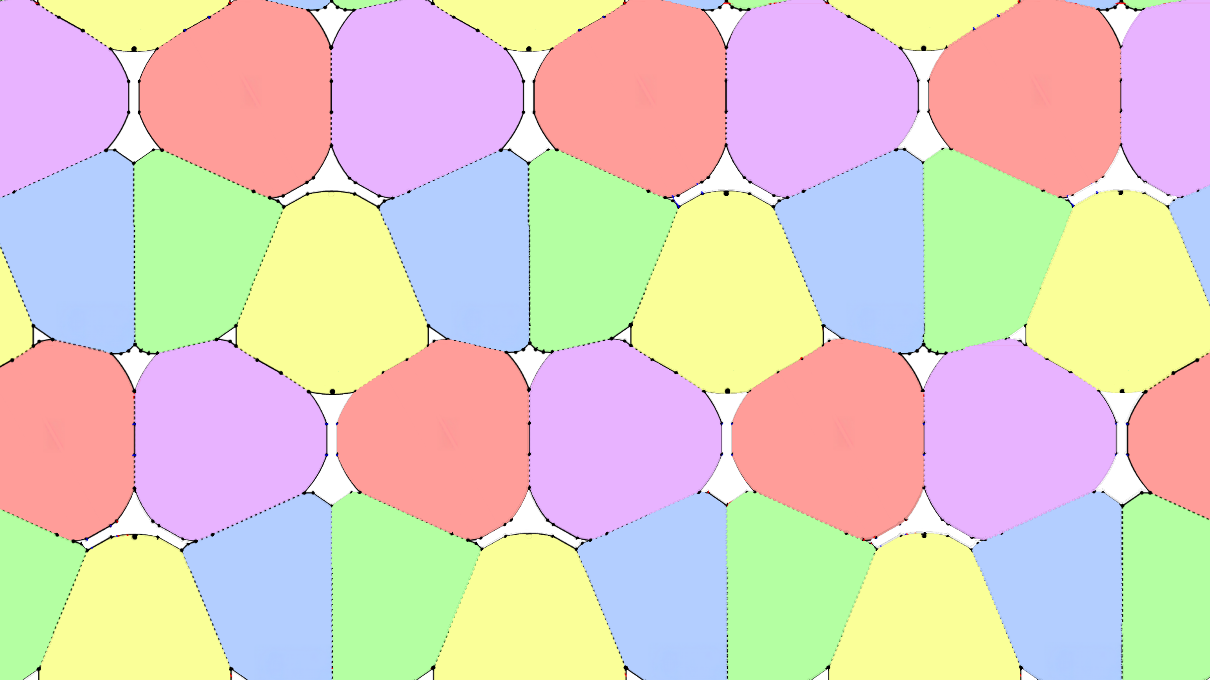
colors	**1**	**2**	**3**	**4**	**5**	**6**
best known	77.06%	54.13%	31.20%	8.25%	4.01%	0.02%
numerics	75.86%	54.14%	31.23%	8.27%	**3.56%**	0.02%

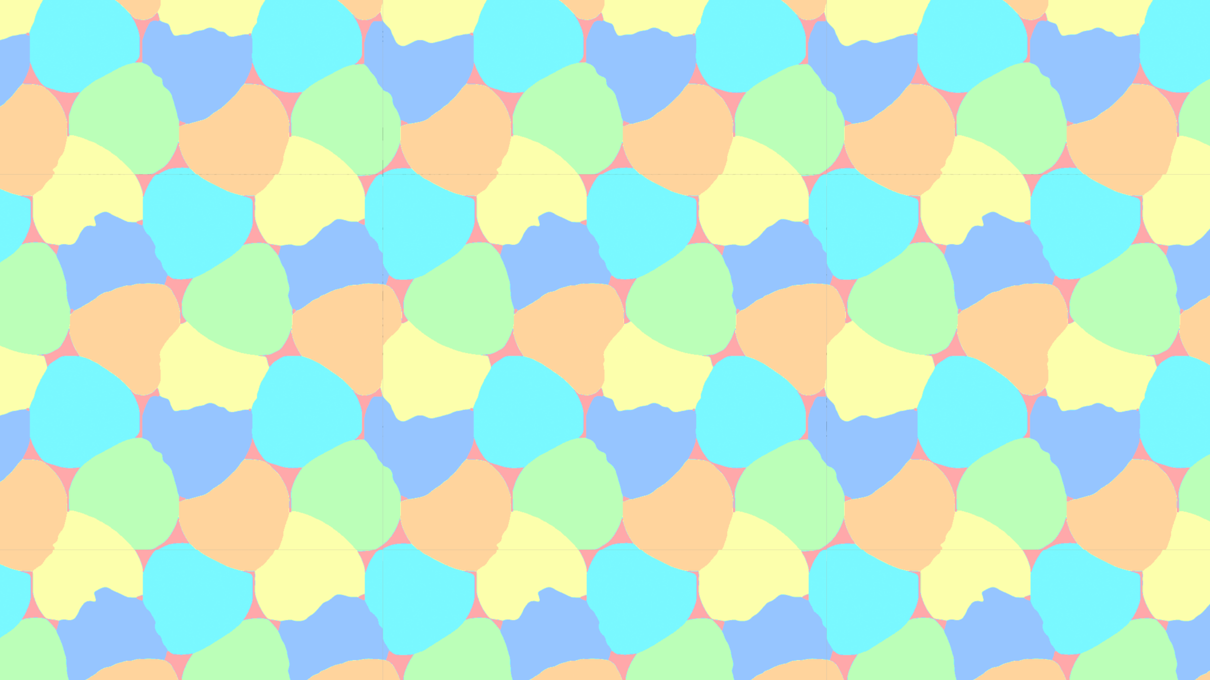
All previously best known values are due to Paarts (2020) building on work of Pritikin (1998) for 6 colors and Croft (1967) for 1, 2, 3, and 4 colors.

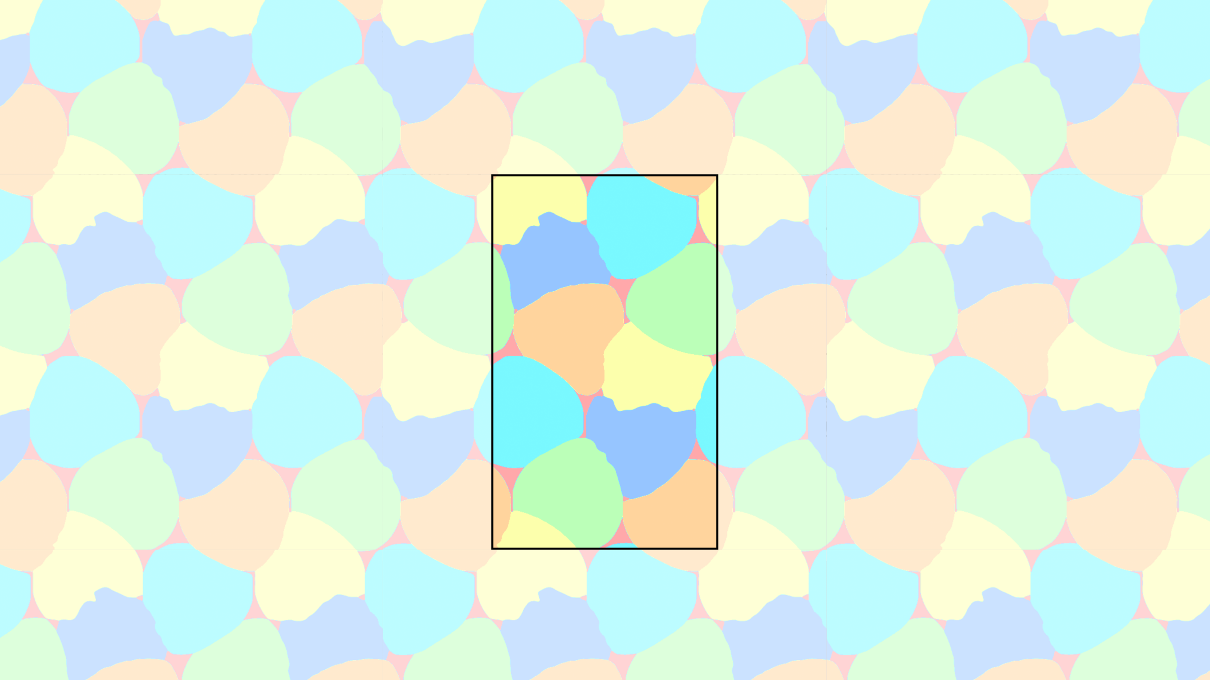
Theorem (Mundinger, Pokutta, S., Zimmer 2025+)

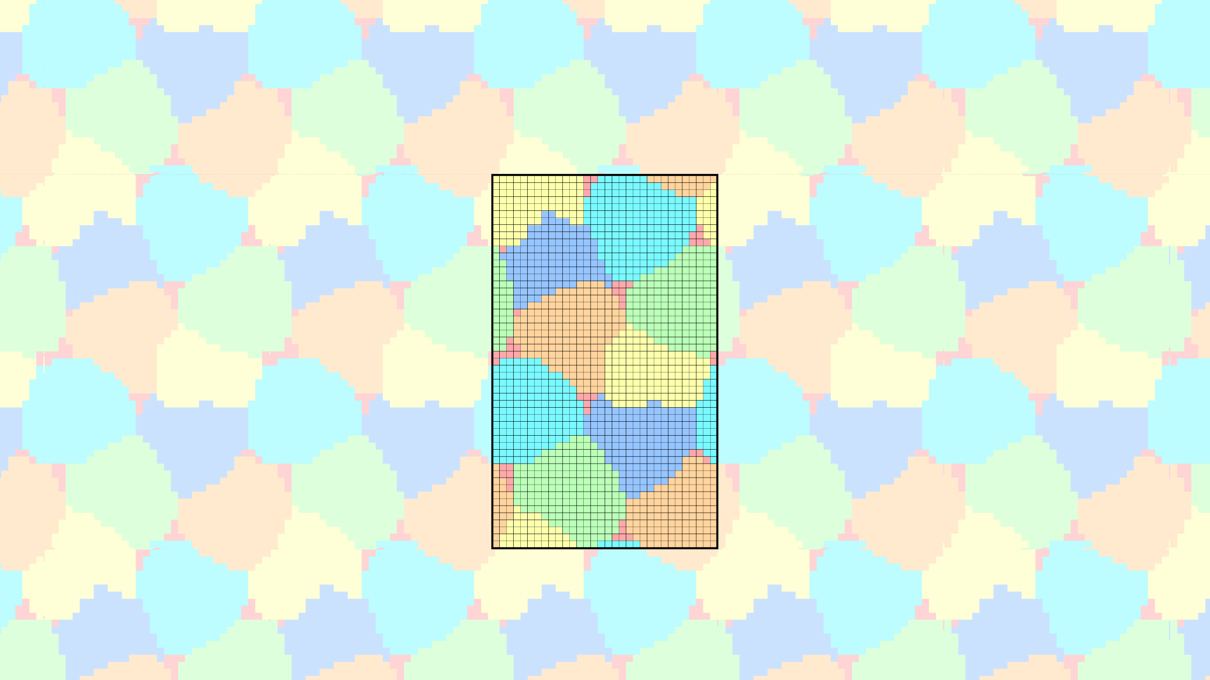
96.29% of the plane can be 5-colored with no monochromatic unit distance pairs.

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	77.06%	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*)

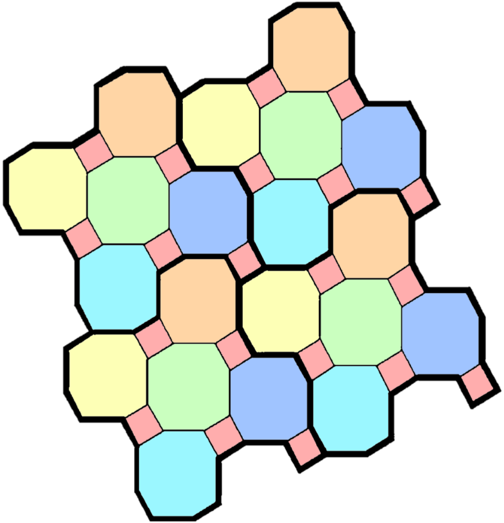
Determine the set of d for which $(1, 1, 1, 1, 1, d)$ can be realized.

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* $(d_1, \dots, d_c)$ if color i does not contain distance d_i.

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

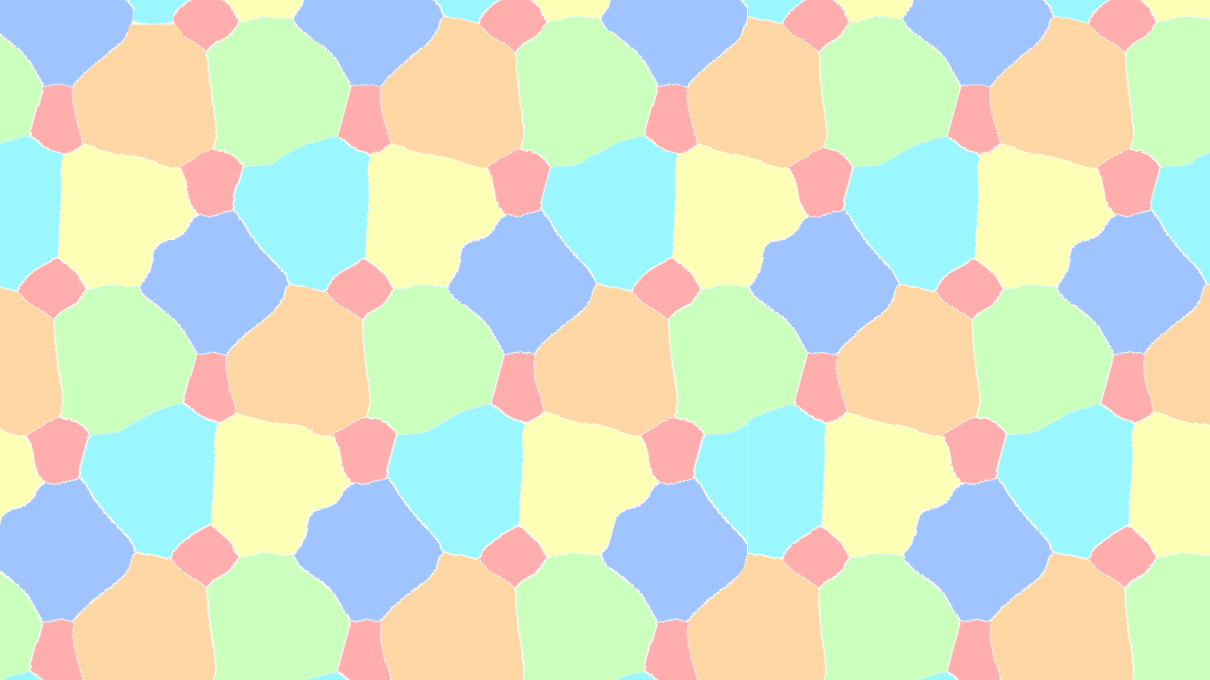
Determine the set of d for which $(1, 1, 1, 1, 1, d)$ *can be realized.*

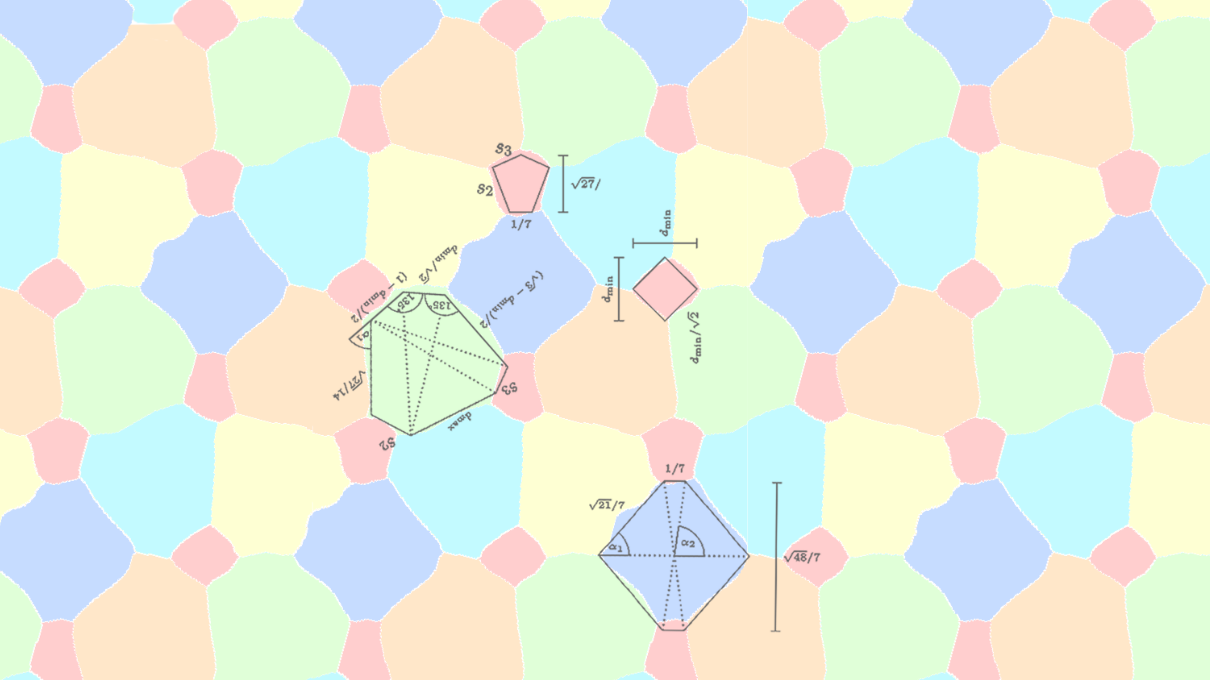
Soifer (1991) found a coloring for $d = 1/\sqrt{5}$. Hoffman and Soifer (1993) also found one for $d = \sqrt{2} - 1$. Both of these are part of a family that covers any

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

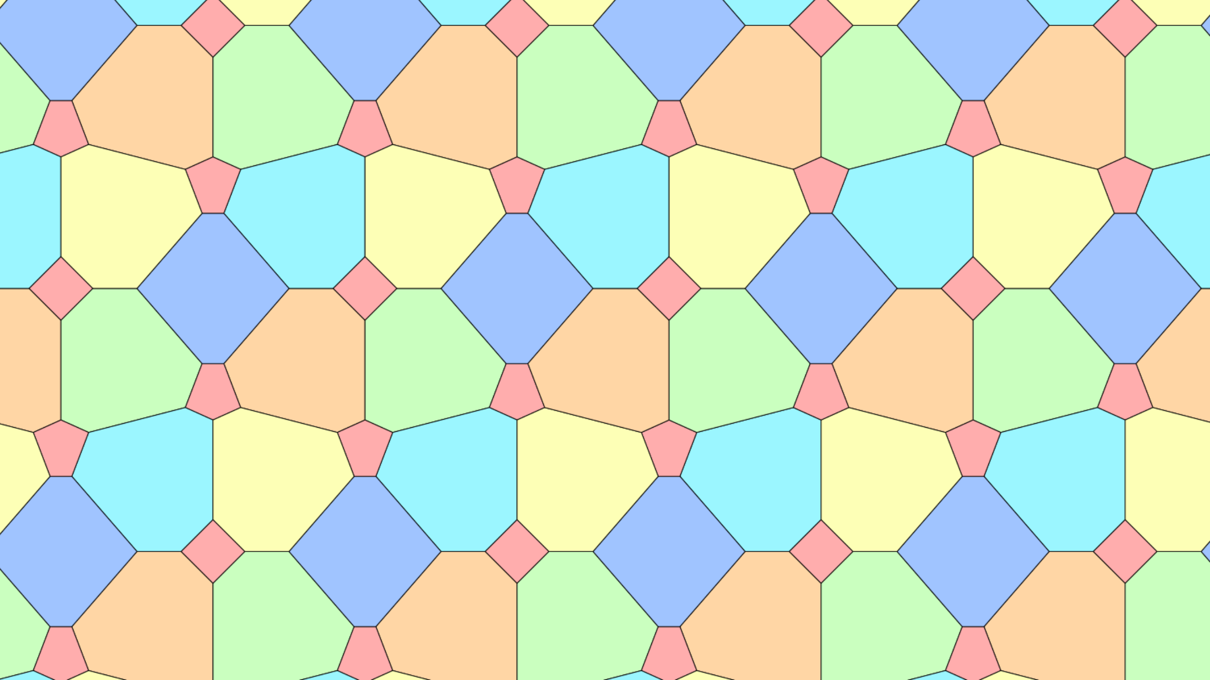
Theorem (Mundinger, Pokutta, S., Zimmer 2024)

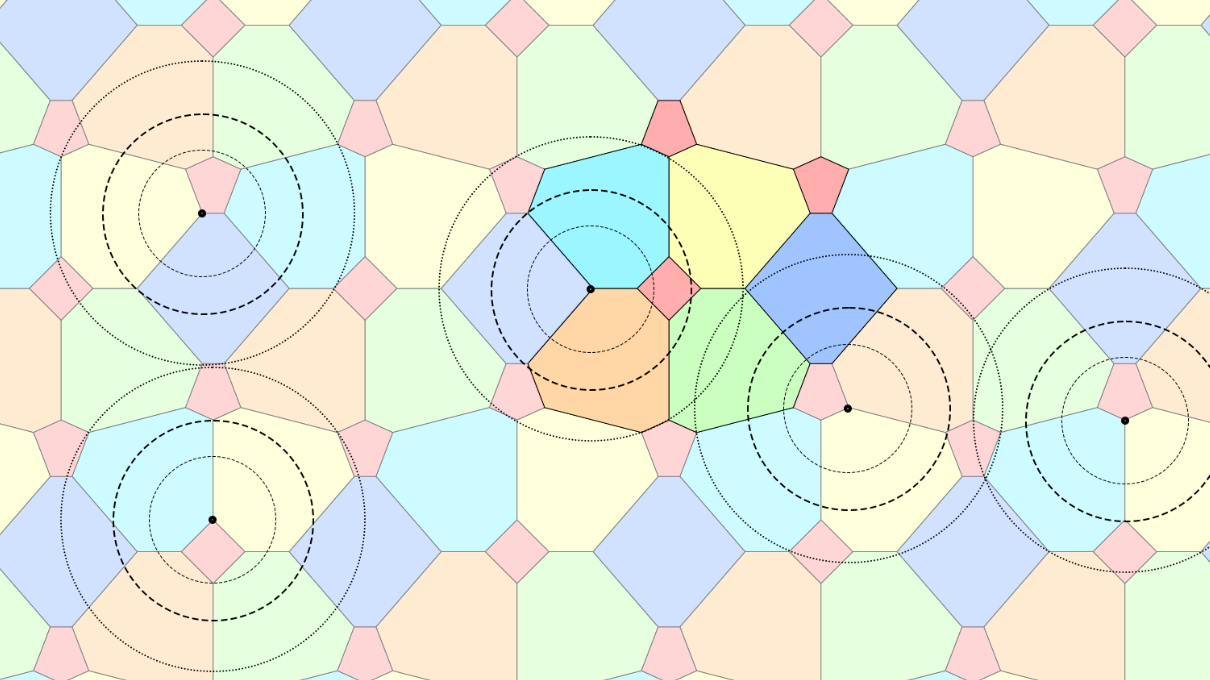
We extended the range of realizable types to $0.354 \leq d \leq 0.553$.

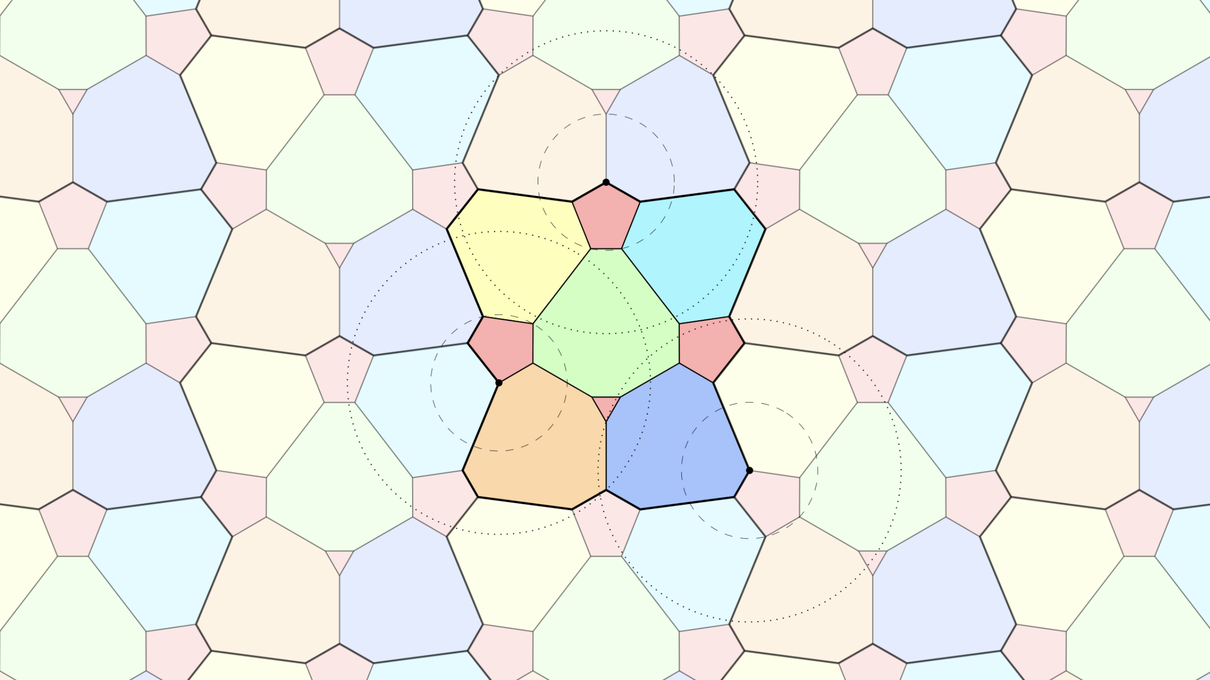




S_3
S_2
$\sqrt{27}/$
1/7
d_{min}
d_{min}
$d_{min}/\sqrt{2}$
$d_{min}/\sqrt{2}$
$(1 - d_{min})/2$
$(\sqrt{3} - d_{min})/2$
135
α_1
$\sqrt{27}/14$
S_3
d_{max}
S_2
1/7
$\sqrt{21}/7$
α_1
α_2
$\sqrt{48}/7$







Variant 2: Going off-diagonal ...

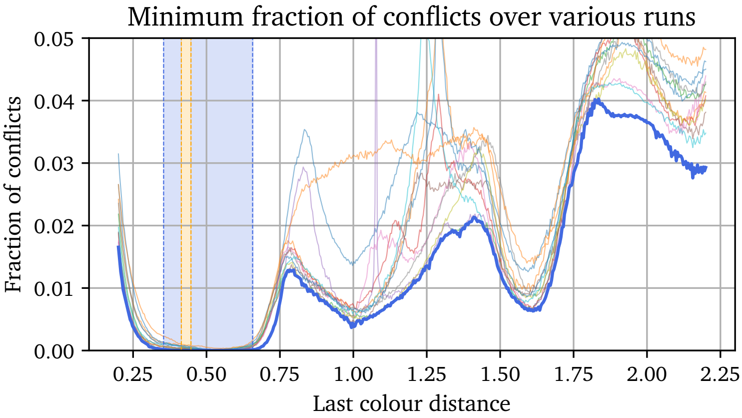
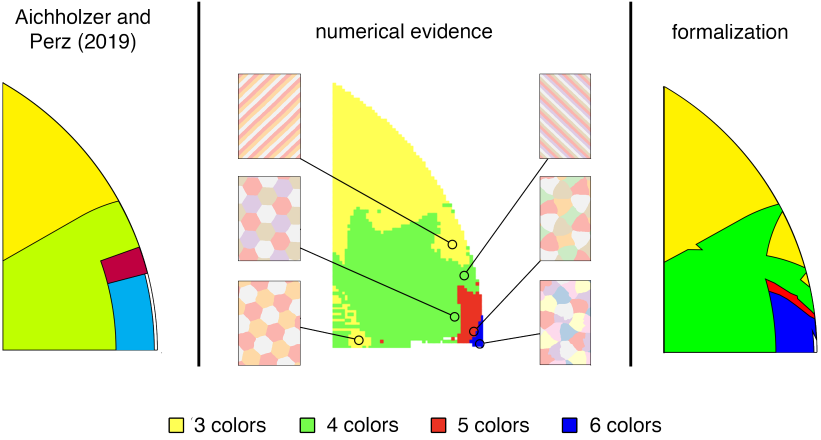


Figure: Numerical results showing the percentage of points with some conflict for a given forbidden distance in the sixth color minimized over several runs.

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 $0 \leq a \leq b \leq 1$?



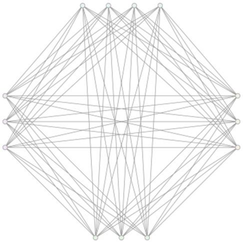
Coloring the Plane with Neural Networks

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.



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

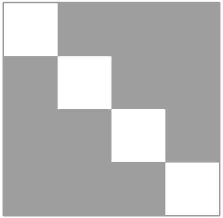


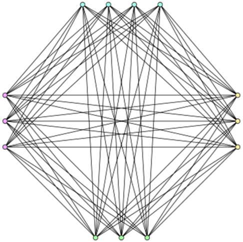
Figure: **(left)** $T(13,4)$ **(center)** adjacency matrix **(right)** graphon of $\left(T(n,4)\right)_{n\in\mathbb{N}}$

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.



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

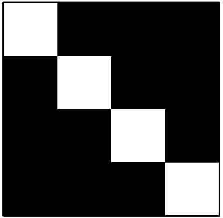


Figure: **(left)** $T(13,4)$ **(center)** adjacency matrix **(right)** graphon of $\big(T(n,4)\big)_{n\in\mathbb{N}}$

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.

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.