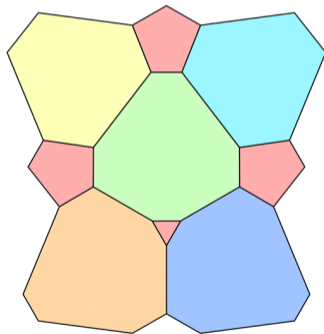


Coloring the Plane with Neural Networks

PACOH Workshop – Jeju Island

K. Mundinger, M. Zimmer, A. Kiem, **C. Spiegel**, S. Pokutta

7th of July 2026



Results are joint work with...



Konrad Mundinger
Zuse Institut Berlin
TU Berlin



Max Zimmer
Zuse Institut Berlin
TU Berlin



Aldo Kiem
Zuse Institut Berlin
TU Berlin



Sebastian Pokutta
Zuse Institut Berlin
TU Berlin



Coloring the Plane with Neural Networks

- 1.** The Chromatic Number of the Plane 4 slides
- 2.** Neural Networks as Universal Approximators 3 slides
- 3.** The Continuum of Six-Colorings 4 slides
- 4.** An Outlook on Other Applications 1 slide
- 5.** Graphons and Graph Limits

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)* ...

1. The Chromatic Number of the Plane

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 ”

The history of the problem

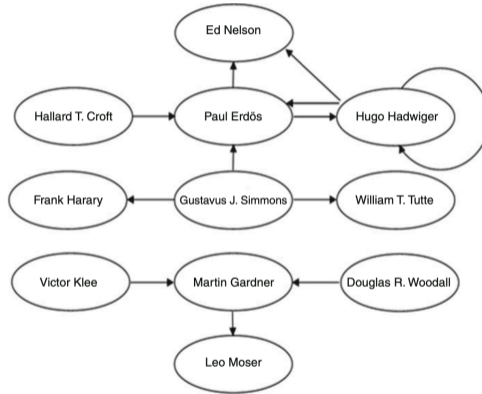


Diagram 3.1 Who created the chromatic number of the plane problem?

1. The Chromatic Number of the Plane

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.

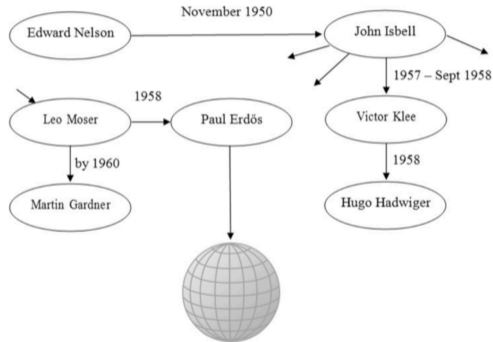


Diagram 3.2 Passing the baton of the chromatic number of the plane problem

Lower bounds through unit distance graphs

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 $f: V \rightarrow \mathbb{E}^2$ of its vertices in the plane s.t. $\|f(u) - f(v)\| = 1$ if and only if $\{u, v\} \in E$.

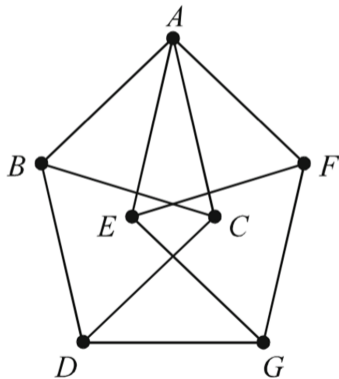
A triangle gives a lower bound of 3 and the Moser spindle a lower bound of 4 (1961).

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

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

Exoo and Ismailescu found a simpler construction with 627 vertices, Heule one with 553 vertices, and Jaan Parts, as part of Polymath16, one with 510 vertices.

Lower bounds through unit distance graphs



Lower bounds through unit distance graphs

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 $f: V \rightarrow \mathbb{E}^2$ 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).

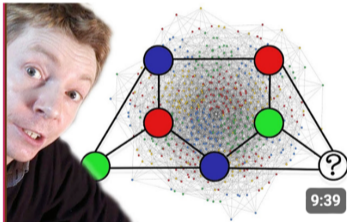
Theorem (Aubrey D.N.J. de Grey, 2018)

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

Exoo and Ismailescu found a simpler construction with 627 vertices, Heule one with 553 vertices, and Jaan Parts, as part of Polymath16, one with 510 vertices.

Lower bounds through unit distance graphs

Numberphile



A Colorful Unsolved Problem - Numberphile

681K views • 5 years ago



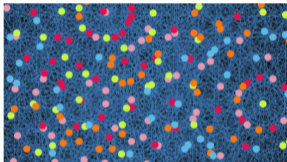
Numberphile ✓

More links & stuff in full description below ↓↓↓

Numberphile is supported by the Mathematical Science...

CC

Lower bounds through unit distance graphs



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..

1. The Chromatic Number of the Plane

Lower bounds through unit distance graphs



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

Lower bounds through unit distance graphs

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 $f: V \rightarrow \mathbb{E}^2$ 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.

Exoo and Ismailescu found a simpler construction with 627 vertices, Heule one with 553 vertices, and Jaan Parts, as part of Polymath16, one with 510 vertices.

Upper bounds through colorings

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

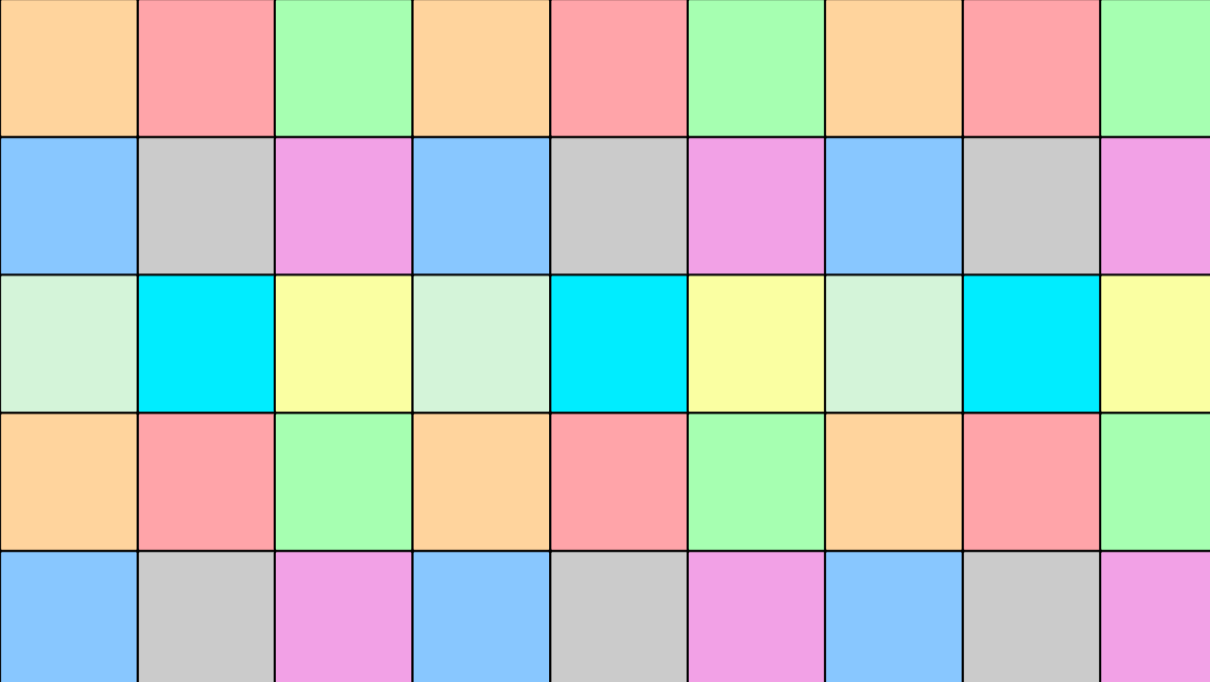
$$5 \leq \chi(\mathbb{E}^2) \leq \dots$$

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

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

Idea. Use a parameterized and easily differentiable family $g_\theta : \mathbb{E}^2 \rightarrow \Delta_c$ and find

$$\arg \min_{\theta} \mathbf{E} \left[\int_{B_1(x)} g_\theta(x) \cdot g_\theta(y) dy \mid x \in \mathbb{E}^2 \right].$$



Upper bounds through colorings

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

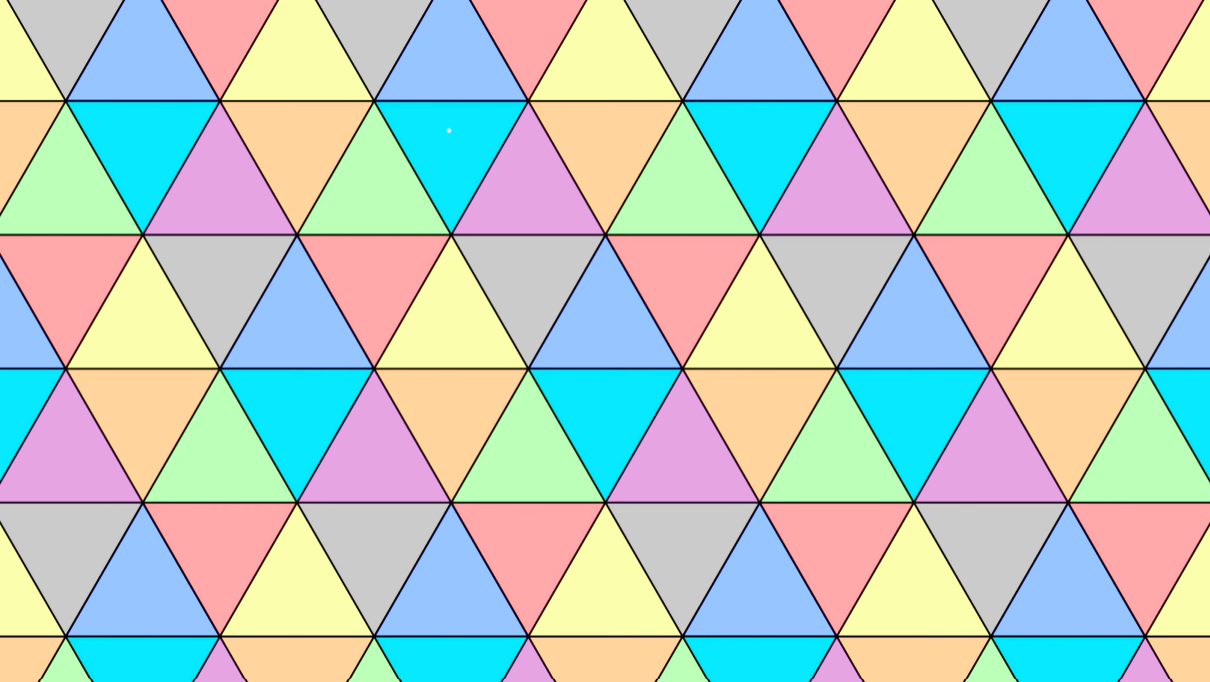
$$5 \leq \chi(\mathbb{E}^2) \leq 9.$$

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

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

Idea. Use a parameterized and easily differentiable family $g_\theta : \mathbb{E}^2 \rightarrow \Delta_c$ and find

$$\arg \min_{\theta} \mathbf{E} \left[\int_{B_1(x)} g_\theta(x) \cdot g_\theta(y) dy \mid x \in \mathbb{E}^2 \right].$$



Upper bounds through colorings

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

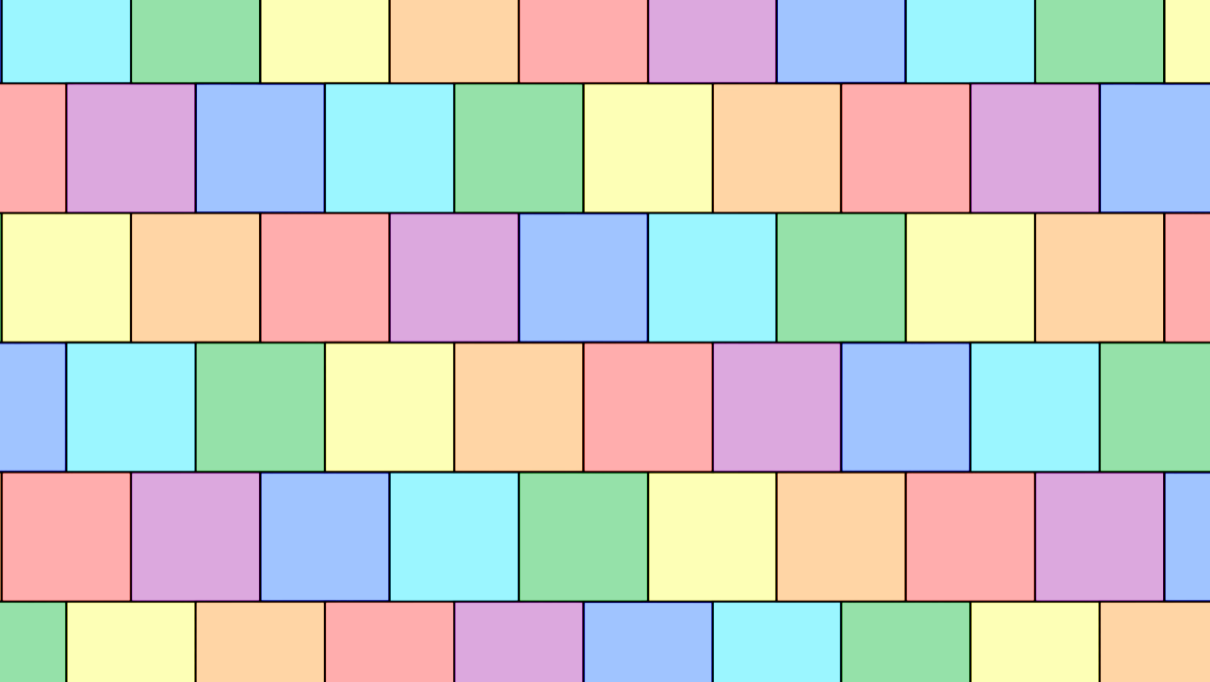
$$5 \leq \chi(\mathbb{E}^2) \leq 8.$$

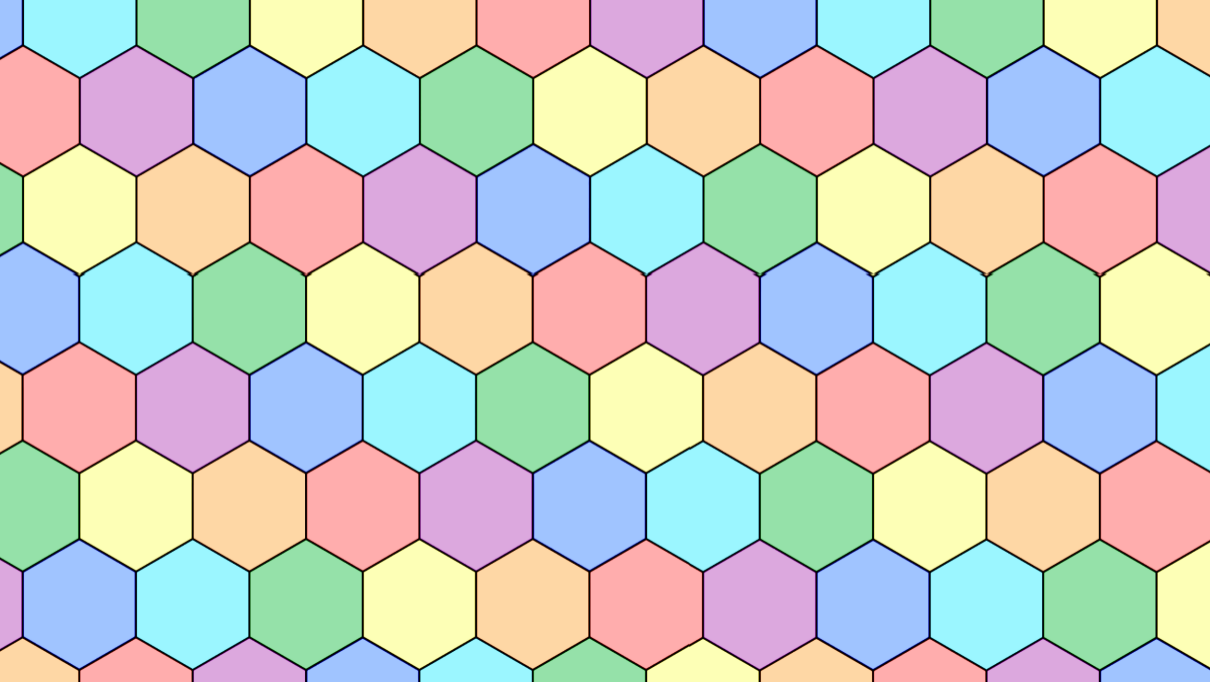
Question. Can we use computers to find colorings $g : \mathbb{E}^2 \rightarrow [c]$ so that

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

Idea. Use a parameterized and easily differentiable family $g_\theta : \mathbb{E}^2 \rightarrow \Delta_c$ and find

$$\arg \min_{\theta} \mathbf{E} \left[\int_{B_1(x)} g_\theta(x) \cdot g_\theta(y) dy \mid x \in \mathbb{E}^2 \right].$$





Upper bounds through colorings

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

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

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

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

Idea. Use a parameterized and easily differentiable family $g_\theta : \mathbb{E}^2 \rightarrow \Delta_c$ and find

$$\arg \min_{\theta} \mathbf{E} \left[\int_{B_1(x)} g_\theta(x) \cdot g_\theta(y) dy \mid x \in \mathbb{E}^2 \right].$$

Upper bounds through colorings

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

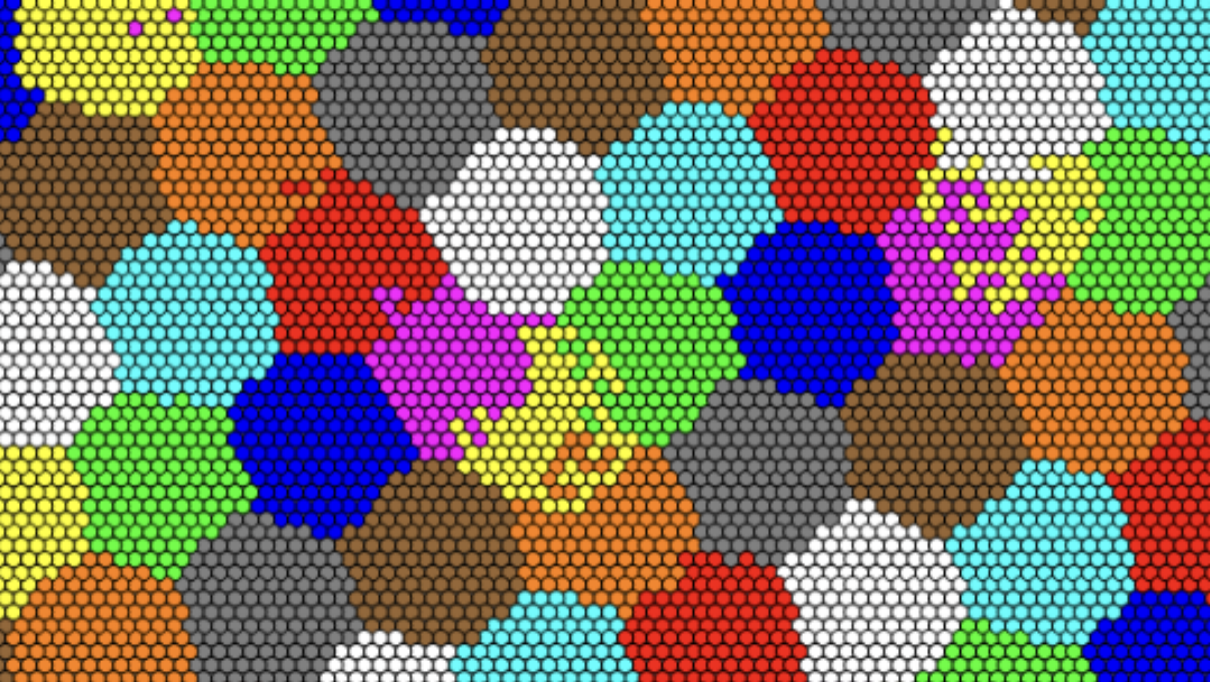
$$5 \leq \chi(\mathbb{E}^2) \leq 7.$$

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

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

Idea. Use a parameterized and easily differentiable family $g_\theta : \mathbb{E}^2 \rightarrow \Delta_c$ and find

$$\arg \min_{\theta} \mathbf{E} \left[\int_{B_1(x)} g_\theta(x) \cdot g_\theta(y) dy \mid x \in \mathbb{E}^2 \right].$$



Upper bounds through colorings

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

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

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

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

Idea. Use a parameterized and easily differentiable family $g_\theta : \mathbb{E}^2 \rightarrow \Delta_c$ and find

$$\arg \min_{\theta} \mathbf{E} \left[\int_{B_1(x)} g_\theta(x) \cdot g_\theta(y) dy \mid x \in \mathbb{E}^2 \right].$$



Coloring the Plane with Neural Networks

1. The Chromatic Number of the Plane 4 slides
2. Neural Networks as Universal Approximators 3 slides
3. The Continuum of Six-Colorings 4 slides
4. An Outlook on Other Applications 1 slide
5. Graphons and Graph Limits

What are Neural Networks?

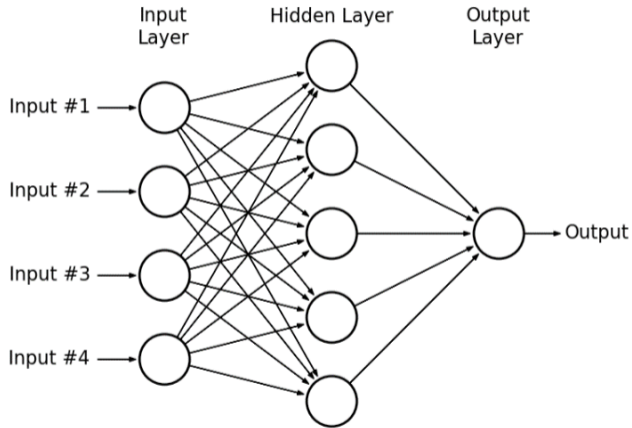


Figure: Feedforward neural network or multilayer perceptron architecture.

2. Neural Networks as Universal Approximators

What are Neural Networks?

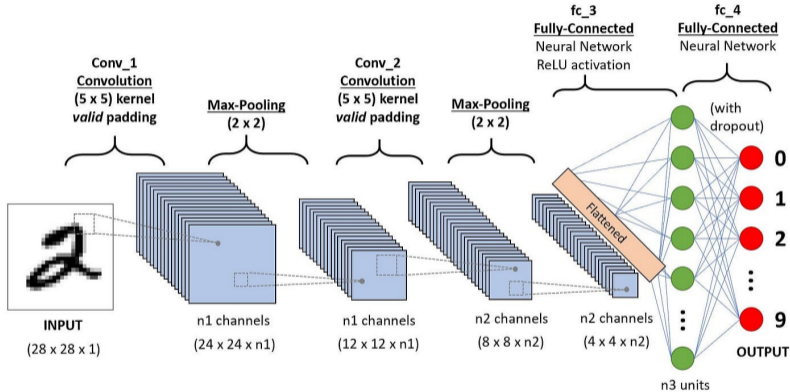


Figure: Convolutional neural network architecture.

What are Neural Networks?

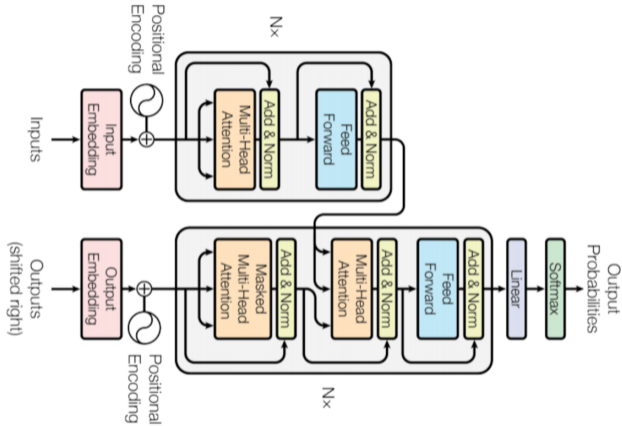
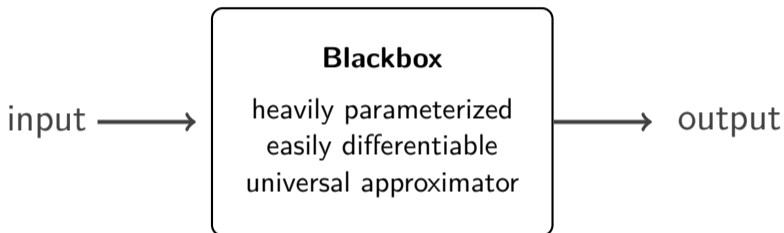


Figure: Transformer neural network architecture.

What are Neural Networks?

Just a parameterized family of functions g_θ with some convenient properties...

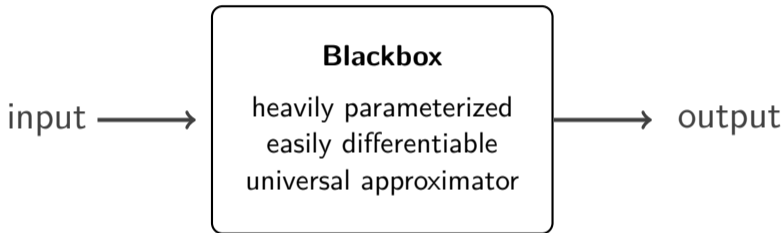


Theorem (Universal Approximation Theorem)

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

What are Neural Networks?

Just a parameterized family of functions g_θ with some convenient properties...



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?

Idea. What if we use batch gradient descent to 'train' $g_\theta : \mathbb{E}^2 \rightarrow \Delta_6$ to minimize

$$L(\theta) = \int_{[-b,b] \times [-b,b]} \int_{B_1(x)} g_\theta(x) \cdot g_\theta(y) dy dx?$$

Algorithm. We sample points $x^{(i)} \in [-b, b] \times [-b, b]$ and $y^{(i)} \in B_1(x)$ and use that

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

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

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

How do we find the correct parameters?

Idea. What if we use batch gradient descent to 'train' $g_\theta : \mathbb{E}^2 \rightarrow \Delta_6$ to minimize

$$L(\theta) = \int_{[-b,b] \times [-b,b]} \int_{B_1(x)} g_\theta(x) \cdot g_\theta(y) dy dx?$$

Algorithm. We sample points $x^{(i)} \in [-b, b] \times [-b, b]$ and $y^{(i)} \in B_1(x)$ and use that

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

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

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

How do we find the correct parameters?

Idea. What if we use batch gradient descent to 'train' $g_\theta : \mathbb{E}^2 \rightarrow \Delta_6$ to minimize

$$L(\theta) = \int_{[-b,b] \times [-b,b]} \int_{B_1(x)} g_\theta(x) \cdot g_\theta(y) dy dx?$$

Algorithm. We sample points $x^{(i)} \in [-b, b] \times [-b, b]$ and $y^{(i)} \in B_1(x)$ and use that

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

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

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



Unfortunately this coloring was already known...

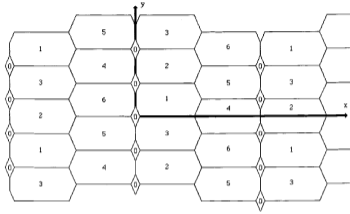
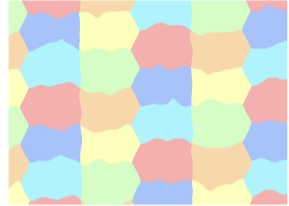
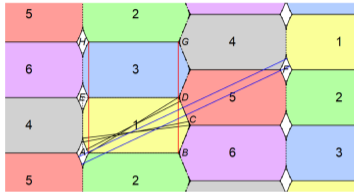


FIG. 3. A good 7-coloring of $(\mathbb{R}^2, 1)$.



Theorem (Pritikin 1995; refined by Parts 2020)

99.985% of the plane can be colored with 6 colors such that no two points of the same color are a unit distance apart.

Corollary

Any unit distance graph with chromatic number 7 must have order at least 6 993.



Coloring the Plane with Neural Networks

1. The Chromatic Number of the Plane 4 slides
2. Neural Networks as Universal Approximators 3 slides
3. The Continuum of Six-Colorings 4 slides
4. An Outlook on Other Applications 1 slide
5. Graphons and Graph Limits

Going off-diagonal...

A variant. A c -coloring of the plane has *coloring type* or *realizes* (d_1, \dots, d_c) if color i does not contain any points at distance d_i .

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

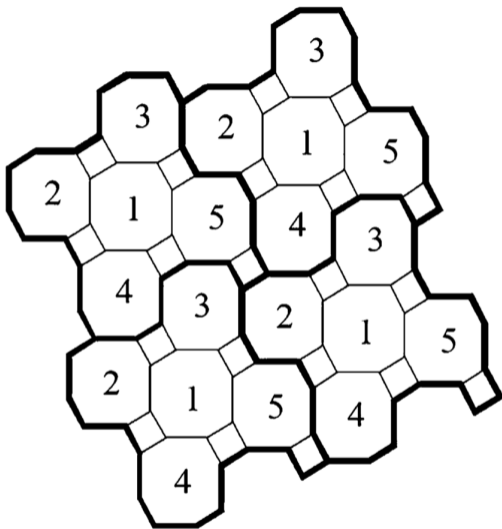
Determine the set of d for which $(1, 1, 1, 1, 1, d)$ can be realized.

Soifer found a coloring for $d = 1/\sqrt{5}$ in 1991. Hoffman and Soifer also found one for $d = \sqrt{2} - 1$ in 1993. 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)

There is a coloring realizing $(1, 1, 1, 1, 1, d)$ for any $0.418 \leq d \leq 0.657$ and another (family of) colorings covers any $0.354 \leq d \leq 0.553$.



Going off-diagonal...

A variant. A c -coloring of the plane has *coloring type* or *realizes* (d_1, \dots, d_c) if color i does not contain any points at distance d_i .

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

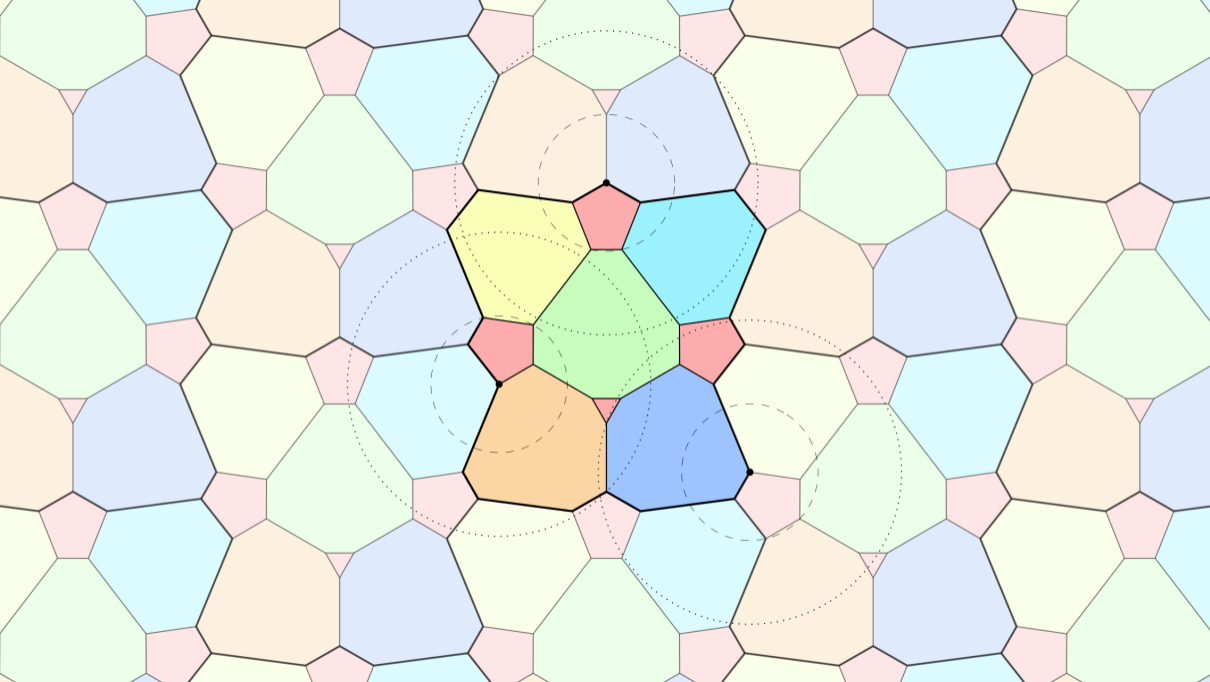
Determine the set of d for which $(1, 1, 1, 1, 1, d)$ can be realized.

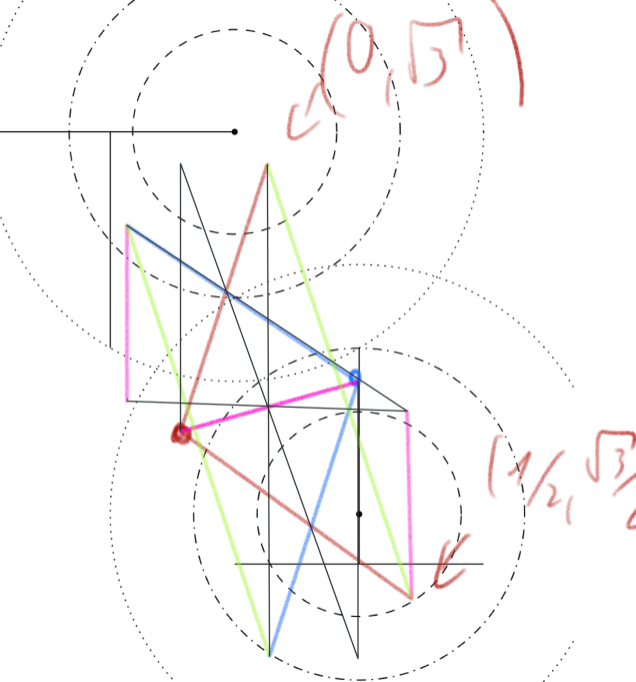
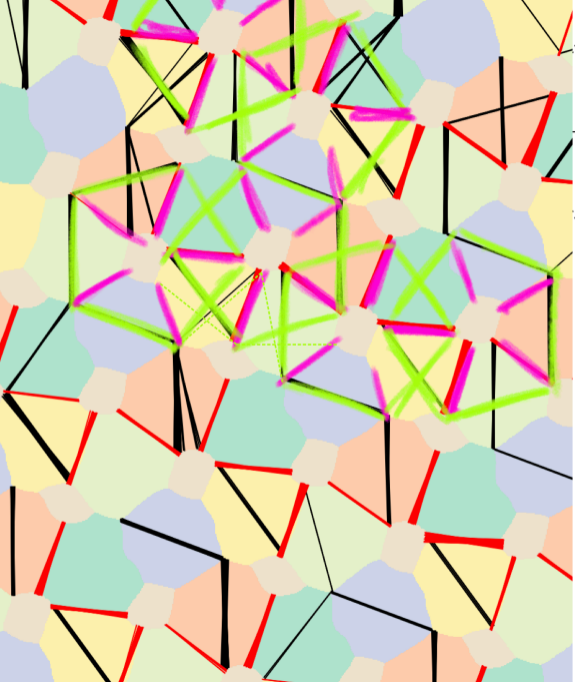
Soifer found a coloring for $d = 1/\sqrt{5}$ in 1991. Hoffman and Soifer also found one for $d = \sqrt{2} - 1$ in 1993. Both of these are part of a family that covers any

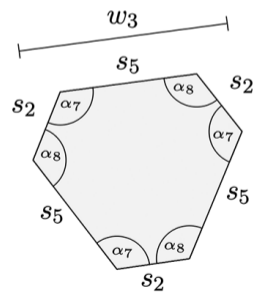
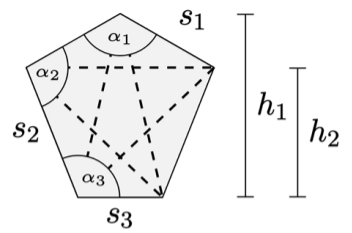
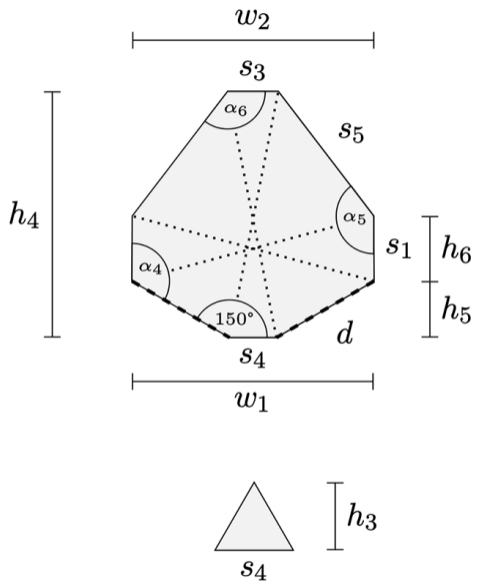
$$0.414 \approx \sqrt{2} - 1 \leq d \leq 1/\sqrt{5} \approx 0.447.$$

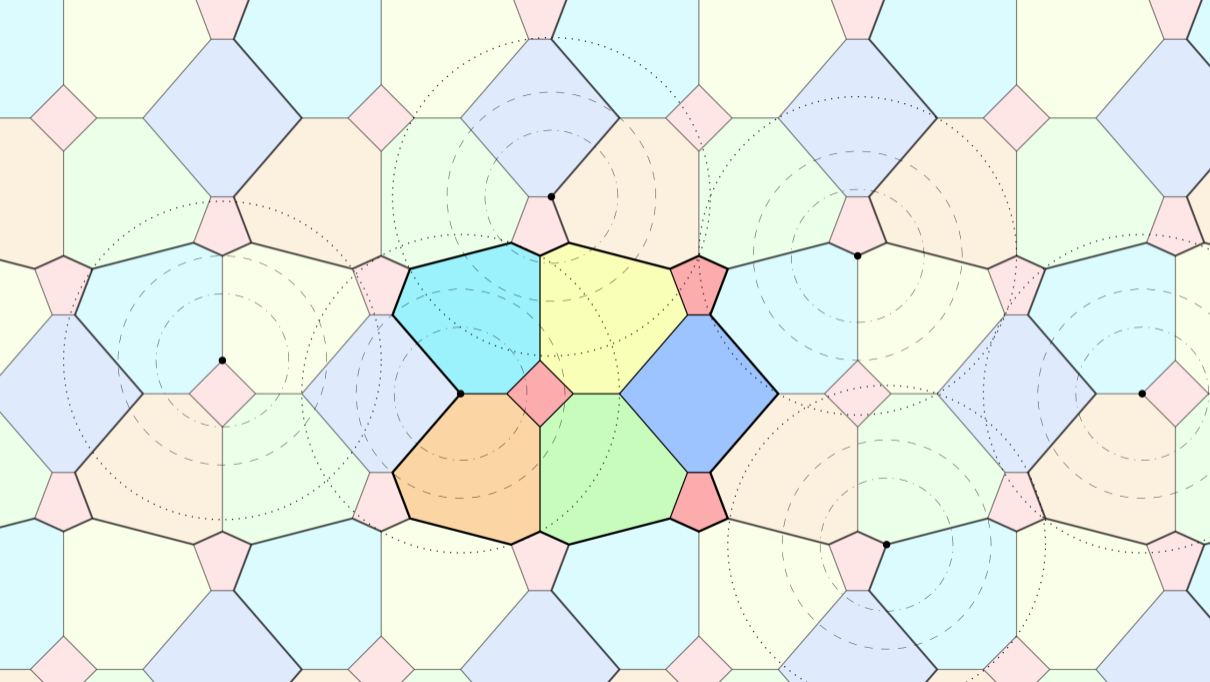
Theorem (Mundinger, Pokutta, S., Zimmer 2024)

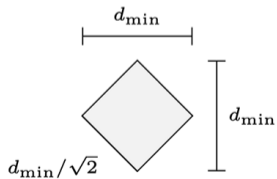
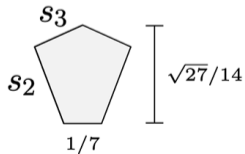
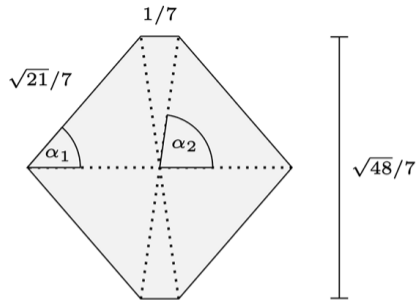
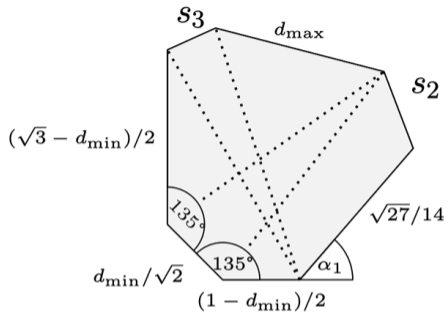
There is a coloring realizing $(1, 1, 1, 1, 1, d)$ for any $0.418 \leq d \leq 0.657$ and another (family of) colorings covers any $0.354 \leq d \leq 0.553$.











Is this optimal?

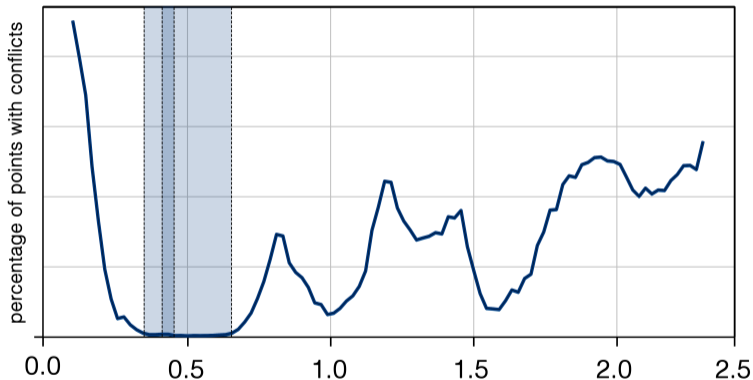


Figure: Numerical results showing the percentage of points with some conflict for a given forbidden distance d in the sixth color found over several minimized over several runs.



Coloring the Plane with Neural Networks

1. The Chromatic Number of the Plane 4 slides
2. Neural Networks as Universal Approximators 3 slides
3. The Continuum of Six-Colorings 4 slides
4. An Outlook on Other Applications 1 slide
5. Graphons and Graph Limits

Open problems and final remarks

The underlying optimization approach is very flexible:

- Can we formalize these colorings as **Voronoi diagrams**?
- Can we improve the upper bound of the **chromatic number of \mathbb{E}^3** from 15 to 14?
- Can we apply the same ideas to generate **graphons and other limit structures**?
- Can we use **adversarial networks** when the objective is non-differentiable?

The *methodology* was accepted at **ICML 2025** (arXiv:2501.18527).

The *colorings* were published in **Geombinatorics Quarterly (2024)** (arXiv:2404.05509).



Coloring the Plane with Neural Networks

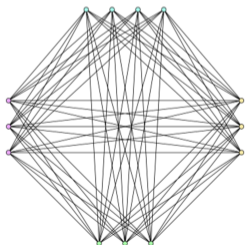
1. The Chromatic Number of the Plane 4 slides
2. Neural Networks as Universal Approximators 3 slides
3. The Continuum of Six-Colorings 4 slides
4. An Outlook on Other Applications 1 slide
5. Graphons and Graph Limits

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 $\mathcal{W} : [0, 1]^2 \rightarrow [0, 1]$ correspond bijectively to convergent graph sequences.

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



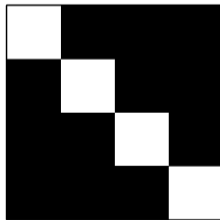
$$\begin{bmatrix} & & & & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ & & & & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ & & & & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ & & & & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & & & & & 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & & & & & 1 & 1 & 1 & 1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


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

Is the same approach applicable?

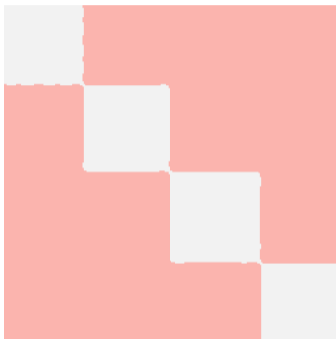


Figure: Result of maximizing the number of edges while penalizing cliques of size 5 with a Lagrangian term.

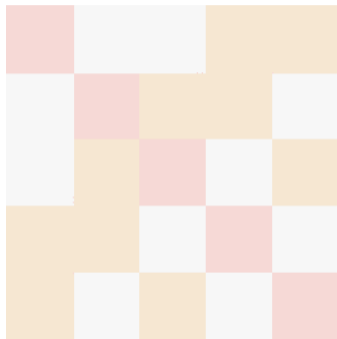


Figure: Result of minimizing the number of monochromatic triangles in 3-colorings of the edges of a complete graph.

Is the same approach applicable?

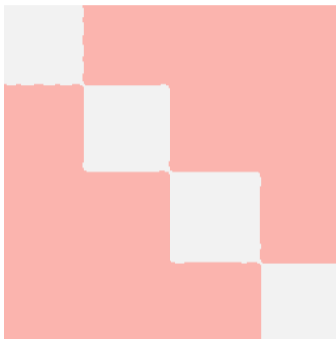


Figure: Result of maximizing the number of edges while penalizing cliques of size 5 with a Lagrangian term.



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

Is the same approach applicable?

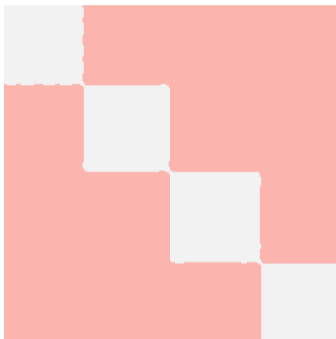


Figure: Result of maximizing the number of edges while penalizing cliques of size 5 with a Lagrangian term.

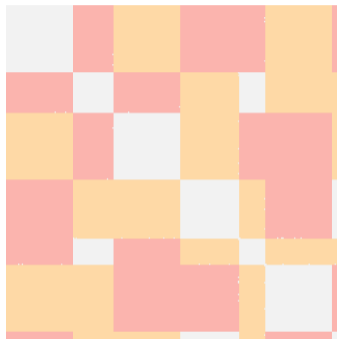


Figure: Result of minimizing the number of monochromatic triangles in 3-colorings of the edges of a complete graph.

Is the same approach applicable?

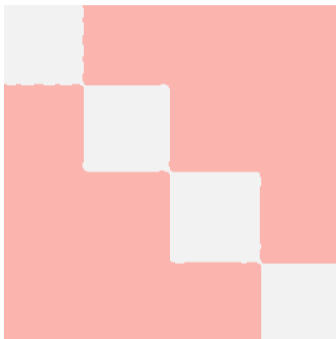


Figure: Result of maximizing the number of edges while penalizing cliques of size 5 with a Lagrangian term.

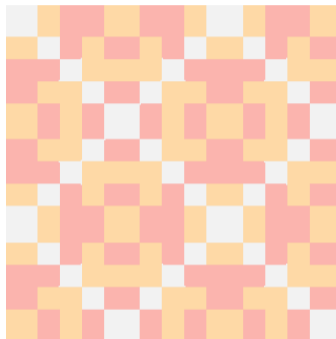


Figure: Result of minimizing the number of monochromatic triangles in 3-colorings of the edges of a complete graph.

Is the same approach applicable?

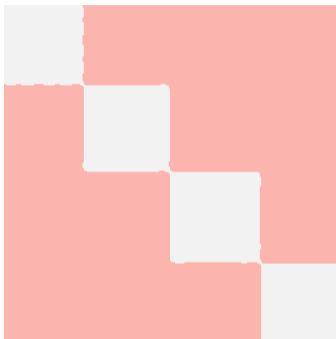


Figure: Result of maximizing the number of edges while penalizing cliques of size 5 with a Lagrangian term.

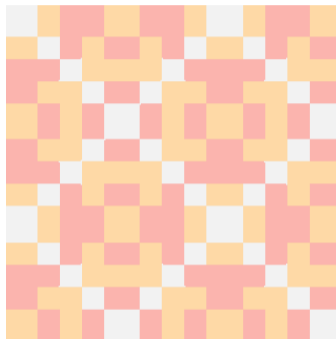


Figure: Result of minimizing the number of monochromatic triangles in 3-colorings of the edges of a complete graph.



Thank you for your attention!