Independence and chromatic number (and random k-SAT): Sparse Case

Dimitris Achlioptas
Microsoft
Random graphs

$G(n, p)$

$G(n, m)$

Each edge appears, independently, with probability $p$.

We add $m$ edges one-by-one.

W.h.p.: with probability that tends to 1 as $n \to \infty$. 

W.h.p.: with probability that tends to 1 as $n \to \infty$. 

Hamiltonian cycle

- Let $\tau_2$ be the moment all vertices have degree $\geq 2$.
Let $\tau_2$ be the moment all vertices have degree $\geq 2$

W.h.p. $G(n, m = \tau_2)$ has a Hamiltonian cycle
Hamiltonian cycle

- Let $\tau_2$ be the moment all vertices have degree $\geq 2$
- W.h.p. $G(n, m = \tau_2)$ has a Hamiltonian cycle
- W.h.p. it can be found in time $O(n^3 \log n)$

[Ajtai, Komlós, Szemerédi 85] [Bollobás, Fenner, Frieze 87]
Hamiltonian cycle

- Let $\tau_2$ be the moment all vertices have degree $\geq 2$
- W.h.p. $G(n, m = \tau_2)$ has a Hamiltonian cycle
- W.h.p. it can be found in time $O(n^3 \log n)$

[Ajtai, Komlós, Szemerédi 85] [Bollobás, Fenner, Frieze 87]

- In $G(n, 1/2)$ Hamiltonicity can be decided in $O(n)$ expected time.

[Gurevich, Shelah 84]
Cliquess in random graphs

- The largest clique in $G(n, 1/2)$ has size
  \[2 \log_2 n - 2 \log_2 \log_2 n \pm 1\]
  
  [Bollobás, Erdős 75] [Matula 76]
Clique in random graphs

- The largest clique in $G(n, 1/2)$ has size
  $$2 \log_2 n - 2 \log_2 \log_2 n \pm 1$$
  [Bollobás, Erdős 75] [Matula 76]

- No maximal clique of size $< \log_2 n$
**Cliques in random graphs**

- The largest clique in $G(n, 1/2)$ has size
  $$2 \log_2 n - 2 \log_2 \log_2 n \pm 1$$
  [Bollobás, Erdős 75] [Matula 76]

- No maximal clique of size $< \log_2 n$

- Can we find a clique of size $(1 + \epsilon) \log_2 n$?  
  [Karp 76]
Cliquies in random graphs

- The largest clique in $G(n, 1/2)$ has size
  \[ 2 \log_2 n - 2 \log_2 \log_2 n \pm 1 \]

  [Bollobás, Erdős 75] [Matula 76]

- No maximal clique of size $< \log_2 n$

- Can we find a clique of size $(1 + \varepsilon) \log_2 n$?

  What if we “hide” a clique of size $n^{1=2i^2}$?
Two problems for which we know much less.

- Chromatic number of sparse random graphs
- Random k-SAT
Two problems for which we know much less.

- Chromatic number of sparse random graphs
- Random k-SAT

Canonical for random constraint satisfaction:
  - Binary constraints over k-ary domain
  - k-ary constraints over binary domain

Studied in: AI, Math, Optimization, Physics, …
A factor-graph representation of k-coloring

- Each **vertex** is a variable with domain \( \{1, 2, \ldots, k\} \).
- Each **edge** is a constraint on two variables.
- All constraints are “not-equal”.
- Random graph = each constraint picks two variables at random.
SAT via factor-graphs

\((\overline{x}_{12} \lor x_{5} \lor \overline{x}_{9}) \land (x_{34} \lor \overline{x}_{21} \lor x_{5}) \land \cdots \cdots \land (x_{21} \lor x_{9} \lor \overline{x}_{13})\)
SAT via factor-graphs

\( \overline{x_{12}} \lor x_5 \lor \overline{x_9} \land (x_{34} \lor \overline{x_{21}} \lor x_5) \land \cdots \land (x_{21} \lor x_9 \lor \overline{x_{13}}) \)

- Edge between \( x \) and \( c \) iff \( x \) occurs in clause \( c \).
- Edges are labeled +/- to indicate whether the literal is negated.
- Constraints are “at least one literal must be satisfied”.
- Random \( k \)-SAT = constraints pick \( k \) literals at random.
Diluted mean-field spin glasses

- Small, discrete domains: *spins*
- Conflicting, fixed constraints: *quenched disorder*
- Random bipartite graph: *lack of geometry, mean field*
- Sparse: *diluted*
- Hypergraph coloring, random XOR-SAT, error-correcting codes…
Random graph coloring: Background
A trivial lower bound

- For any graph, the chromatic number is at least:

  \[
  \frac{\text{Number of vertices}}{\text{Size of maximum independent set}}
  \]
A trivial lower bound

- For any graph, the chromatic number is at least:
  \[
  \frac{\text{Number of vertices}}{\text{Size of maximum independent set}}
  \]

- For random graphs, use upper bound for largest independent set.
  \[
  \frac{\mu \prod_{s}^{n} (1 - \rho)^{\binom{s}{2}}}{s} \to 0
  \]
An algorithmic upper bound

- Repeat
  - Pick a random uncolored vertex
  - Assign it the lowest allowed number (color)

Uses $2 \times$ trivial lower bound number of colors
An algorithmic upper bound

- Repeat
  - Pick a random uncolored vertex
  - Assign it the lowest allowed number (color)

Uses 2 x trivial lower bound number of colors

- No algorithm is known to do better
The lower bound is asymptotically tight

As \( d \) grows, \( G(n, d/n) \) can be colored using independent sets of essentially maximum size

[Bollobás 89]
[Łuczak 91]
The lower bound is asymptotically tight

As $d$ grows, $G(n, d/n)$ can be colored using independent sets of essentially maximum size

[Bollobás 89]
[Łuczak 91]

<table>
<thead>
<tr>
<th>Average degree</th>
<th>$10^{60}$</th>
<th>$10^{80}$</th>
<th>$10^{100}$</th>
<th>$10^{130}$</th>
<th>$10^{1000}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lower bound</td>
<td>$37 \cdot 10^{56}$</td>
<td>$28 \cdot 10^{76}$</td>
<td>$22 \cdot 10^{96}$</td>
<td>$17 \cdot 10^{126}$</td>
<td>$21 \cdot 10^{995}$</td>
</tr>
<tr>
<td>Upper / Lower</td>
<td>1.97</td>
<td>1.78</td>
<td>1.68</td>
<td>1.53</td>
<td>1.14</td>
</tr>
</tbody>
</table>
Theorem. For every $d > 0$, there exists an integer $k = k(d)$ such that w.h.p. the chromatic number of $G(n, p = d/n)$ is either $k$ or $k + 1$

[Łuczak 91]
Theorem. For every $d > 0$, there exists an integer $k = k(d)$ such that w.h.p. the chromatic number of $G(n, p = d/n)$ is either $k$ or $k + 1$

where $k$ is the smallest integer s.t. $d < 2k \log k$. 

“The Values”
Examples

- If $d = 7$, w.h.p. the chromatic number is 4 or 5.
Examples

- If $d = 7$, w.h.p. the chromatic number is 4 or 5.

- If $d = 10^{60}$, w.h.p. the chromatic number is
  
  377145549067226075809014239493833600551612641764765068157

  or

  3771455490672260758090142394938336005516126417647650681576
Theorem. If \((2k - 1) \ln k < d < 2k \ln k\) then w.h.p. the chromatic number of \(G(n, d/n)\) is \(k + 1\).
Theorem. If \((2k - 1) \ln k < d < 2k \ln k\) then w.h.p. the chromatic number of \(G(n, d/n)\) is \(k + 1\).

- If \(d = 10^{100}\) then w.h.p. the chromatic number is
Random k-SAT: Background
Random $k$-SAT

- Fix a set of $n$ variables $\mathcal{X} = \{x_1, x_2, \ldots, x_n\}$
- Among all $\binom{n}{k}$ possible $k$-clauses select $m$ uniformly and independently. Typically $m = rn$.

Example ($k = 3$):

$$(\overline{x}_{12} \lor x_5 \lor \overline{x}_9) \land (x_{34} \lor \overline{x}_{21} \lor x_5) \land \cdots \land (x_{21} \lor x_9 \lor \overline{x}_{13})$$
Generating hard 3-SAT instances

[Mitchell, Selman, Levesque 92]
Generating hard 3-SAT instances

[Mitchell, Selman, Levesque 92]

- The critical point appears to be around $r \approx 4.2$
The satisfiability threshold conjecture

- For every $k \geq 3$, there is a constant $r_k$ such that

$$\lim_{n \to \infty} \Pr[\mathcal{F}_k(n, r n) \text{ is satisfiable}] = \begin{cases} \frac{1}{2} & \text{if } r = r_k - \epsilon \\ 1 & \text{if } r = r_k \\ 0 & \text{if } r = r_k + \epsilon \end{cases}$$
The satisfiability threshold conjecture

- For every $k \geq 3$, there is a constant $r_k$ such that

$$\lim_{n \to \infty} \Pr[\mathcal{F}_k(n, rn) \text{ is satisfiable}] = \begin{cases} \frac{1}{2} & \text{if } r = r_k - \epsilon \\ 1 & \text{if } r = r_k \\ 0 & \text{if } r = r_k + \epsilon \end{cases}$$

- For every $k \geq 3$,

$$\frac{2^k}{k} < r_k < 2^k \ln 2$$
Unit-clause propagation

Repeat
– Pick a random unset variable and set it to 1
– While there are unit-clauses satisfy them
– If a 0-clause is generated fail
Unit-clause propagation

Repeat

- Pick a random unset variable and set it to 1
- While there are unit-clauses satisfy them
- If a 0-clause is generated fail

UC finds a satisfying truth assignment if

\[ r < \frac{2^k}{k} \]

[Chao, Franco 86]
An asymptotic gap

- The probability of satisfiability is at most

\[
\frac{\mu}{2^n} \left(1 - \frac{1}{2^k}\right)^m \cdot \frac{\mu}{2} \left(1 - \frac{1}{2^k}\right)^{r, n} \rightarrow 0 \quad \text{for } r \geq 2^k \ln 2
\]
An asymptotic gap

Since mid-80s, no asymptotic progress over

$$\frac{2^k}{\kappa} < r_k < 2^k \ln 2$$
Getting to within a factor of 2

**Theorem:** For all $\kappa \geq 3$ and

$$r < 2^{k^i} \ln 2 - 1$$

a random $\kappa$-CNF formula with $m = r n$ clauses w.h.p. has a complementary pair of satisfying truth assignments.
The trivial upper bound is the truth!

**Theorem:** For all $k \geq 3$, a random $k$-CNF formula with $m = rn$ clauses is w.h.p. satisfiable if

$$r \leq 2^k \ln 2 - \frac{k}{2} - 1$$
Some explicit bounds for the k-SAT threshold

<table>
<thead>
<tr>
<th>$k$</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>7</th>
<th>20</th>
<th>21</th>
</tr>
</thead>
<tbody>
<tr>
<td>Upper bound</td>
<td>4.51</td>
<td>10.23</td>
<td>21.33</td>
<td>87.88</td>
<td>726,817</td>
<td>1,453,635</td>
</tr>
<tr>
<td>Our lower bound</td>
<td>2.68</td>
<td>7.91</td>
<td>18.79</td>
<td>84.82</td>
<td>726,809</td>
<td>1,453,626</td>
</tr>
<tr>
<td>Algorithmic lower bound</td>
<td>3.52</td>
<td>5.54</td>
<td>9.63</td>
<td>33.23</td>
<td>95,263</td>
<td>181,453</td>
</tr>
</tbody>
</table>
For any non-negative r.v. $X$, 

$$ \Pr[X > 0] \geq \frac{\mathbb{E}[X]^2}{\mathbb{E}[X^2]} $$

**Proof:** Let $Y = 1$ if $X > 0$, and $Y = 0$ otherwise. By Cauchy-Schwartz,

$$ \mathbb{E}[X]^2 = \mathbb{E}[XY]^2 \leq \mathbb{E}[X^2]\mathbb{E}[Y^2] = \mathbb{E}[X^2]\Pr[X > 0] $$.
Ideal for sums

If $X = X_1 + X_2 + \cdots$ then

$$E[X]^2 = E[X_i]E[X_j]$$

$$E[X^2] = E[X_i X_j]$$
Ideal for sums

If \( X = X_1 + X_2 + \cdots \) then

\[
E[X]^2 = \sum_{i,j} E[X_i]E[X_j]
\]

\[
E[X^2] = \sum_{i,j} E[X_iX_j]
\]

Example:

The \( X_i \) correspond to the potential \( q \)-cliques in \( G(n, 1/2) \). Dominant contribution from non-overlapping cliques.
General observations

- Method works well when the $X_i$ are like “needles in a haystack”
General observations

- Method works well when the $X_i$ are like "needles in a haystack"

- Lack of correlations $\implies$ rapid drop in influence around solutions
General observations

- Method works well when the $X_i$ are like “needles in a haystack”
- Lack of correlations $\implies$ rapid drop in influence around solutions
- Algorithms get no “hints”
The second moment method for random k-SAT

- Let $X$ be the number of satisfying truth assignments
The second moment method for random $k$-SAT

- Let $X$ be the # of satisfying truth assignments

For every clause-density $r > 0$, there is $\beta = \beta(r) > 0$ such that

$$\frac{\mathbb{E}[X^2]}{\mathbb{E}[X^2]} < (1 - \beta)^n$$
The second moment method for random k-SAT

- Let $X$ be the number of satisfying truth assignments.

For every clause-density $r > 0$, there is $\beta = \beta(r) > 0$ such that

\[
\frac{E[X^2]}{E[X^2]} < (1 - \beta)^n
\]

- The number of satisfying truth assignments has huge variance.
The second moment method for random k-SAT

- Let $X$ be the # of satisfying truth assignments.

  For every clause-density $r > 0$, there is $\beta = \beta(r) > 0$ such that
  \[
  \frac{E[X^2]}{E[X^2]} < (1 - \beta)^n
  \]

- The number of satisfying truth assignments has huge variance.

- The satisfying truth assignments do not form a “uniformly random mist” in $\{0, 1\}^n$. 
To prove $2^k \ln 2 - k/2 - 1$

- Let $H(\sigma, F)$ be the number of satisfied literal occurrences in $F$ under $\sigma$
To prove $2^k \ln 2 - k/2 - 1$

- Let $H(\sigma, F)$ be the number of satisfied literal occurrences in $F$ under $\sigma$

- Let $X = X(F)$ be defined as

\[
X(F) = \sum_{\gamma} 1_{\gamma 

\begin{align*}
&X_{\gamma}^H(\frac{3}{4}F) \\
&X_{\gamma}^{\frac{3}{4}} Y
\end{align*}
\]

where $\gamma < 1$ satisfies $(1 + \gamma^2)^k i \quad (1 - \gamma^2) = 1$. 
To prove $2^k \ln 2 - k/2 - 1$

- Let $H(\sigma, F)$ be the number of satisfied literal occurrences in $F$ under $\sigma$.

- Let $X = X(F)$ be defined as

$$X(F) = X^{\frac{3}{4}} Y^{\frac{1}{4}} F^{\gamma^H(\frac{3}{4}F)}$$

$$X^{\frac{3}{4}} Y^{\frac{1}{4}} = X^{\frac{3}{4}} c^{\gamma^H(\frac{3}{4}c)}$$

where $\gamma < 1$ satisfies $(1 + \gamma^2)^k i^1 (1 - \gamma^2) = 1$. 

- $\gamma < 1$ satisfies $(1 + \gamma^2)^k i^1 (1 - \gamma^2) = 1$. 


General functions

- Given any t.a. $\sigma$ and any $k$-clause $c$ let

  \[ v = v(\sigma, c) \in \{-1, +1\}^k \]

  be the values of the literals in $c$ under $\sigma$. 
General functions

- Given any t.a. $\sigma$ and any $k$-clause $c$ let

  \[ v = v(\sigma, c) \in \{-1, +1\}^k \]

  be the values of the literals in $c$ under $\sigma$.

- We will study random variables of the form

  \[ X = \prod_{y} f(v(\sigma, c)) \prod_{y} c \]

  where $f : \{-1, +1\}^k \to \mathbb{R}$ is an arbitrary function
\[ X = X \times Y \]

\[ f(v(\sigma, c)) \]

\[ \frac{3}{4} \]

\[ c \]

\[ f(v) = 1 \quad \text{for all } v \]

\[ \Rightarrow \quad 2^n \]

\[ f(v) = 0 \quad \text{if } v = (-1, -1, \ldots, -1) \]

\[ \Rightarrow \quad \# \text{ of satisfying truth assignments} \]

\[ f(v) = 1 \quad \text{otherwise} \]

\[ \Rightarrow \quad \# \text{ of } \text{"Not All Equal" truth assignments (NAE)} \]

\[ f(v) = 0 \quad \text{if } v = (-1, -1, \ldots, -1) \text{ or } \quad \Rightarrow \]

\[ f(v) = 1 \quad \text{if } v = (+1, +1, \ldots, +1) \]

\[ \Rightarrow \quad \text{otherwise} \]
\[ X = \frac{3}{4} + c \]

\[ f(v(\sigma, c)) \]

<table>
<thead>
<tr>
<th>( f(v) )</th>
<th>=</th>
<th>1 for all ( v )</th>
<th>( \Rightarrow )</th>
<th>( 2^n )</th>
</tr>
</thead>
<tbody>
<tr>
<td>( f(v) = 0 )</td>
<td>if ( v = (-1, -1, \ldots, -1) )</td>
<td>( \Rightarrow )</td>
<td># of satisfying truth assignments</td>
<td></td>
</tr>
<tr>
<td>( f(v) = 1 )</td>
<td>otherwise</td>
<td>( \Rightarrow )</td>
<td># of satisfying truth assignments whose complement is also satisfying</td>
<td></td>
</tr>
</tbody>
</table>

\[ f(v) = \begin{cases} 
0 & \text{if } v = (-1, -1, \ldots, -1) \\
1 & \text{otherwise} \\
8 & \text{if } v = (-1, -1, \ldots, -1) \text{ or } +1, +1, \ldots, +1 \\
1 & \text{otherwise}
\end{cases} \]
Overlap parameter = distance

- Overlap parameter is Hamming distance
Overlap parameter = distance

- Overlap parameter is Hamming distance

- For any $f$, if $\sigma$, $\tau$ agree on $z = n/2$ variables

\[
\begin{align*}
\mathbb{E} f(v(\sigma, c)) f(v(\tau, c)) &= \mathbb{E} f(v(\sigma, c)) \mathbb{E} f(v(\tau, c))
\end{align*}
\]
Overlap parameter = distance

- Overlap parameter is Hamming distance

- For any $f$, if $\sigma, \tau$ agree on $\frac{z}{2} = \frac{n}{2}$ variables

$$
\mathbb{E} f(v(\sigma, c))f(v(\tau, c)) = \mathbb{E} \mathbb{E} f(v(\sigma, c)) \mathbb{E} f(v(\tau, c))
$$

- For any $f$, if $\sigma, \tau$ agree on $\frac{z}{n}$ variables, let

$$
C_f(\frac{z}{n}) \equiv \mathbb{E} f(v(\sigma, c))f(v(\tau, c))
$$
Contribution according to distance

\[ E[X^2] = \sum_{\sigma, \tau} \prod_c E[f(\sigma, c)f(\tau, c)] \]

\[ = \sum_{\sigma, \tau} \left( E[f(\sigma, c)f(\tau, c)] \right)^m \]  

\[ = 2^n \sum_{z=0}^{n} \binom{n}{z} C_f(z/n)^m \]  

Independence

Identically distributed

Fixing \( \sigma \)
Entropy vs. correlation

For every function $f$:

\[
E [X^2] = 2^n \sum_{z=0}^{\mu_n} \prod_{z} C_f (z/n)^m
\]

\[
E [X]^2 = 2^n \sum_{z=0}^{\mu_n} \prod_{z} C_f (1/2)^m
\]
Contribution according to distance

\[ E[X^2] = \sum_{\sigma, \tau} \prod_c E[f(\sigma, c)f(\tau, c)] \]

\[ = \sum_{\sigma, \tau} \left( E[f(\sigma, c)f(\tau, c)] \right)^m \]

\[ = 2^n \sum_{z=0}^n \binom{n}{z} C_f(z/n)^m \]

\[ = \left( \max_{0 \leq \alpha \leq 1} \frac{2 \alpha^r}{\alpha^\alpha (1 - \alpha)^{1-\alpha}} \right)^n \times \Theta(1) \]

Independence

Identically distributed

Fixing \( \sigma \)

Laplace method
\[
\frac{1}{\alpha^\alpha (1 - \alpha)^{1-\alpha}}
\]

\[\mathcal{C}_{SAT}(\alpha) = 1 - 2 \cdot 2^{-k} + 2^{-k} \alpha^k\]

\[\circled{z/n}\]
The importance of being balanced

- An analytic condition:

\[ C_f(1=2) \neq 0 \implies \text{the s.m.m. fails} \]
\[ \frac{1}{\alpha^\alpha (1 - \alpha)^{1 - \alpha}} \]

\[ \mathcal{C}_{NAE}(\alpha) = 1 - 2^{-k+2} + \frac{\alpha^k + (1 - \alpha)^k}{2^{k-1}} \]

\[ \text{\(\mathcal{C} \equiv z/n\)} \]
NAE 5-SAT

$7 < r < 11$
NAE 5-SAT

$7 < r < 11$
The importance of being balanced

- An **analytic** condition:

\[ C_f(1=2) \neq 0 \implies \text{the s.m.m. fails} \]
The importance of being balanced

- **An analytic condition:**

  \[ C_f(1=2) \neq 0 \iff \text{the s.m.m. fails} \]

- **A geometric criterion:**

  \[ C_f(1=2) = 0 \iff f(v)v = 0 \]

  \[ v \in \{-1;+1\}^k \]
The importance of being balanced

\[ C_f(1=2) = 0 \iff f(v)v = 0 \]

\[ v \in \{-1, +1\}^k \]

Constant

SAT

Complementary
Balance & Information Theory

- Want to balance vectors in “optimal” way
Want to balance vectors in “optimal” way

Information theory

maximize the entropy of the $f(v)$ subject to

\[ f(-1, -1, \ldots, -1) = 0 \quad \text{and} \quad f(v)\mathbf{v} = 0 \]

\[ \mathbf{v}^T 2f_i 1;+1 g^k \]
Balance & Information Theory

- Want to balance vectors in “optimal” way
- Information theory

maximize the entropy of the \( f(\mathbf{v}) \) subject to

\[
\begin{align*}
  f(-1, -1, \ldots, -1) &= 0 \\
  \mathbf{v}^T g &= 0
\end{align*}
\]

- Lagrange multipliers

the optimal \( f \) is

\[
f(\mathbf{v}) = \gamma \text{ # of +1s in } \mathbf{v}
\]

for the unique \( \gamma \) that satisfies the constraints
Balance & Information Theory

- Want to balance vectors in “optimal” way
- Information theory

 maximize the entropy of the $f(v)$ subject to

\[
f(-1, -1, \ldots, -1) = 0 \quad \text{and} \quad f(v)v = 0
\]

$\mathbf{X}$

Heractic

((-1, -1, \ldots, -1) and (1, 1, \ldots, 1))
Random graph coloring
Threshold formulation

**Theorem.** A random graph with $n$ vertices and $m = cn$ edges is w.h.p. $k$-colorable if

$$c \leq k \log k - \log k - 1$$

and w.h.p. non-$k$-colorable if

$$c \geq k \log k - \frac{1}{2} \log k.$$
Main points

- Non-$k$-colorability:

**Proof.** The probability that there exists any $k$-coloring is at most

$$\mu^n \prod_{cn} \left(1 - \frac{1}{k}\right) \to 0$$

- $k$-colorability:

**Proof.** Apply second moment method to the number of balanced $k$-colorings of $G(n, m)$. 

Setup

• Let $X_\sigma$ be the indicator that the balanced $k$-partition $\sigma$ is a proper $k$-coloring.

• We will prove that if $X = \frac{3}{4} X^{3/4}$ then for all $c \leq k \log k - \log k - 1$ there is a constant $D = D(k)$ such that

\[ \mathbb{E}[X^2] < D \mathbb{E}[X]^2 \]

• This implies that $G(n, cn)$ is $k$-colorable w.h.p.
Setup

- \( E[X^2] = \text{sum over all } \sigma, \tau \text{ of } E[X_{\sigma}X_{\tau}]. \)

- For any pair of balanced k-partitions \( \sigma, \tau \)
  let \( a_{ij} \) be the # of vertices having color \( i \) in \( \sigma \) and color \( j \) in \( \tau \).

\[
\Pr[\sigma \text{ and } \tau \text{ are proper}] = 1 \frac{1}{cn} - \frac{2}{k} + \sum_{ij} a_{ij}^2 A
\]
Examples

Balance $\implies A$ is doubly-stochastic.

When $\sigma, \tau$ are uncorrelated, $A$ is the flat $1/k$ matrix.
Examples

Balance $\Rightarrow A$ is doubly-stochastic.

When $\sigma, \tau$ are uncorrelated, $A$ is the flat $1/k$ matrix

As $\sigma, \tau$ align, $A$ tends to the identity matrix $I$
A matrix-valued overlap

So, \( E[X^2] = \sum_{a \in A} \tau_{\mu n} - \frac{2}{k} + \frac{1}{k^2} \sum_{a \in A} \tau_{\mu n}^2 \triangleq \bar{n} \)
A matrix-valued overlap

So, \( \mathbb{E}[X^2] = \sum_{A \in B_k} \mu^A \mathbb{E}[X^2] \)

which is controlled by the maximizer of

\[
\sum_{i,j} a_{ij} \log a_{ij} + c \log \left(1 - \frac{2}{k} + \frac{1}{k^2}\right) \sum_{i,j} a_{ij}^2
\]

over \( k \times k \) doubly-stochastic matrices \( A = (a_{ij}) \).
A matrix-valued overlap

So, \( \mathbb{E}[X^2] = \sum_{A,B} \mu^n \mu_{An} 1 - \frac{2}{k} + \frac{1}{k^2} \sum_{a_{ij}} \frac{1}{k^2} a_{ij}^2 \)

which is controlled by the maximizer of

\[
\sum_{a_{ij}} a_{ij} \log a_{ij} + c
\]

over \( k \times k \) doubly-stochastic matrices \( A = (a_{ij}) \).
Entropy decreases away from the flat $\frac{1}{k}$ matrix

\[
\sum_{i,j} a_{ij} \log a_{ij} + c \cdot \frac{1}{k^2} \sum_{i,j} a_{ij}^2
\]
Entropy decreases away from the flat $1/k$ matrix

- For small $c$, this loss overwhelms the sum of squares gain
Entropy decreases away from the flat $1/k$ matrix

- For small $c$, this loss overwhelms the sum of squares gain
- But for large enough $c$,....
Entropy decreases away from the flat $\frac{1}{k}$ matrix

- For small $c$, this loss overwhelms the sum of squares gain
- But for large enough $c$, .....
Entropy decreases away from the flat $1/k$ matrix

- For small $c$, this loss overwhelms the sum of squares gain
- But for large enough $c$, ....

The maximizer jumps instantaneously from flat, to a matrix where $k$ entries capture majority of mass

This jump happens only after $c > k \log k - \log k - 1$
Proof. Compare the value at the flat matrix with upper bound for everywhere else derived by:
Proof overview

Proof. Compare the value at the flat matrix with upper bound for everywhere else derived by:

1. Relax to singly stochastic matrices.
Proof overview

Proof. Compare the value at the flat matrix with upper bound for everywhere else derived by:

1. Relax to singly stochastic matrices.
2. Prescribe the $L_2$ norm of each row $\rho_i$. 
Proof. Compare the value at the flat matrix with upper bound for everywhere else derived by:

1. Relax to singly stochastic matrices.
2. Prescribe the $L_2$ norm of each row $\rho_i$.
3. Find max-entropy, $f(\rho_i)$, of each row given $\rho_i$. 
Proof overview

Proof. Compare the value at the flat matrix with upper bound for everywhere else derived by:

1. Relax to singly stochastic matrices.
2. Prescribe the $L_2$ norm of each row $\rho_i$.
3. Find max-entropy, $f(\rho_i)$, of each row given $\rho_i$.
4. Prove that $f''' > 0$. 
Proof overview

Proof. Compare the value at the flat matrix with upper bound for everywhere else derived by:

1. Relax to singly stochastic matrices.
2. Prescribe the L₂ norm of each row \( \rho_i \).
3. Find max-entropy, \( f(\rho_i) \), of each row given \( \rho_i \).
4. Prove that \( f''' > 0 \).
5. Use (4) to determine the optimal distribution of the \( \rho_i \) given their total \( \rho \).
Proof overview

Proof. Compare the value at the flat matrix with upper bound for everywhere else derived by:

1. Relax to singly stochastic matrices.
2. Prescribe the L$_2$ norm of each row $\rho_i$.
3. Find max-entropy, $f(\rho_i)$, of each row given $\rho_i$.
4. Prove that $f''' > 0$.
5. Use (4) to determine the optimal distribution of the $\rho_i$ given their total $\rho$.
6. Optimize over $\rho$. 
Theorem. For every integer $d > 0$, w.h.p. the chromatic number of a random $d$-regular graph is either $k$, $k + 1$, or $k + 2$ where $k$ is the smallest integer s.t. $d < 2k \log k$. 
A vector analogue (optimizing a single row)

Maximize

\[
X^k - \sum_{i=1}^{X} a_i \log a_i
\]

subject to

\[
X^k - \sum_{i=1}^{X} a_i = 1
\]

\[
X^k - \sum_{i=1}^{X} a_i^2 = \rho
\]

for some \( \frac{1}{k} < \rho < 1 \)
Maximize
\[ \chi^k \]
\[ - \sum_{i=1}^{k} a_i \log a_i \]
subject to
\[ \sum_{i=1}^{k} a_i = 1 \]
\[ \sum_{i=1}^{k} a_i^2 = \rho \]
for some \( 1/k < \rho < 1 \)
A vector analogue (optimizing a single row)

Maximize
\[
\sum_{i=1}^{k} a_i \log a_i
\]
subject to
\[
\sum_{i=1}^{k} a_i = 1
\]
\[
\sum_{i=1}^{k} a_i^2 = \rho
\]
for some \(1/k < \rho < 1\)

For \(k = 3\) the maximizer is \((x, y, y)\) where \(x > y\)
Maximize
\[ X^k \]
\[ - \sum_{i=1}^{k} a_i \log a_i \]
subject to
\[ X^k \]
\[ a_i = 1 \]
\[ i=1 \]
\[ X^k \]
\[ a_i^2 = \rho \]
\[ i=1 \]
for some \( 1/k < \rho < 1 \)

For \( k = 3 \) the maximizer is \((x, y, y)\) where \( x > y \)
Maximize
\[
\sum_{i=1}^{k} a_i \log a_i
\]
subject to
\[
\sum_{i=1}^{k} a_i = 1
\]
\[
\sum_{i=1}^{k} a_i^2 = \rho
\]
for some \( 1/k < \rho < 1 \)

For \( k = 3 \) the maximizer is \((x, y, y)\) where \( x > y \)
Maximize \[
\sum_{i=1}^{k} a_i \log a_i
\]
subject to
\[
\sum_{i=1}^{k} a_i = 1
\]
\[
\sum_{i=1}^{k} a_i^2 = \rho
\]
for some \(\frac{1}{k} < \rho < 1\)

For \(k = 3\) the maximizer is \((x, y, y)\) where \(x > y\)
A vector analogue (optimizing a single row)

Maximize
\[
X^k \quad \text{subject to}
\]
\[
\sum_{i=1}^{k} x_i = 1
\]
\[
\sum_{i=1}^{k} a_i^2 = \rho
\]
for some \( \frac{1}{k} < \rho < 1 \)

For \( k = 3 \) the maximizer is \((x, y, y)\) where \( x > y \)

For \( k > 3 \) the maximizer is \((x, y, \ldots, y)\)
Maximum entropy image restoration

- Create a composite image of an object that:
  - Minimizes “empirical error”
    - Typically, least-squares error over luminance
  - Maximizes “plausibility”
    - Typically, maximum entropy
Maximum entropy image restoration

Structure of maximizer helps detect stars in astronomy
The End