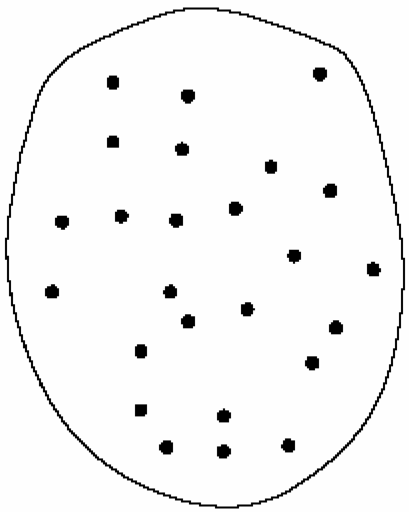
Independence and chromatic number (and random k-SAT): Sparse Case

Dimitris Achlioptas
Microsoft

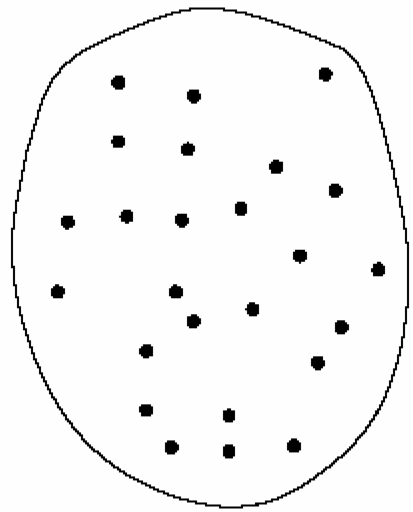
Random graphs

$G(n, p)$



Each edge appears, independently, with probability p.

$G(n, m)$



We add m edges one-by-one.

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

Hamiltonian cycle

- Let τ_2 be the moment all vertices have degree ≥ 2

Hamiltonian cycle

- Let τ_2 be the moment all vertices have degree ≥ 2
  - W.h.p. $G(n, m = \tau_2)$ has a Hamiltonian cycle

Hamiltonian cycle

- Let τ_2 be the moment all vertices have degree ≥ 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 τ_2 be the moment all vertices have degree ≥ 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]

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]

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 < $\log_2 n$

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]

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 $(1+\epsilon)\log_2 n$?

What if we “hide” a clique of size $n^{1/2-\epsilon}$?

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.

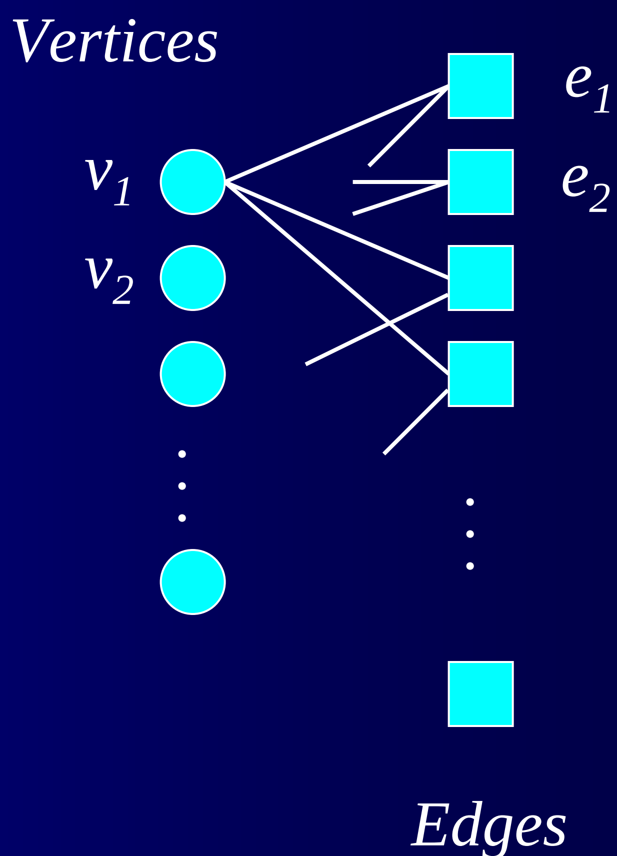
- Chromatic number of sparse random graphs
- Random k-SAT

- 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.



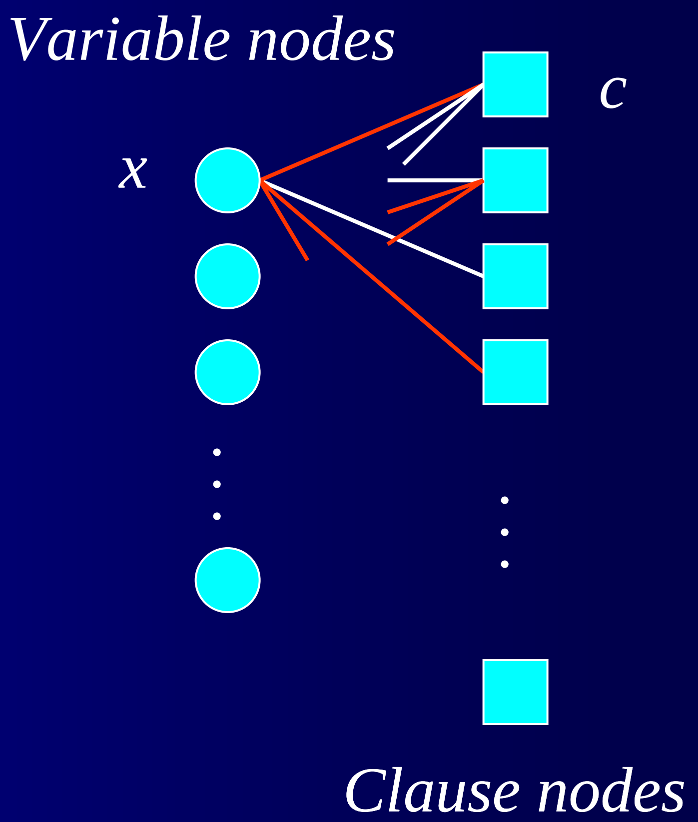
SAT via factor-graphs

$$(\overline{x}_{12} \vee x_5 \vee \overline{x}_9) \wedge (x_{34} \vee \overline{x}_{21} \vee x_5) \wedge \cdots\cdots \wedge (x_{21} \vee x_9 \vee \overline{x}_{13})$$

SAT via factor-graphs

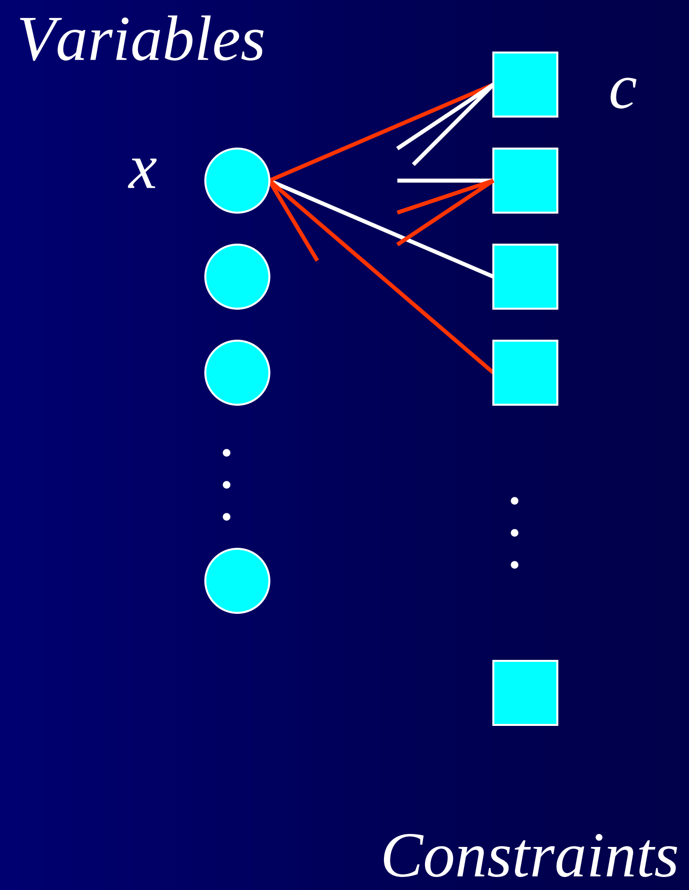
$$(\overline{x}_{12} \vee x_5 \vee \overline{x}_9) \wedge (x_{34} \vee \overline{x}_{21} \vee x_5) \wedge \cdots\cdots \wedge (x_{21} \vee x_9 \vee \overline{x}_{13})$$

- 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.

$$\binom{n}{s} \times (1-p)^{\binom{s}{2}} \to 0$$

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

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]

Average degree	10^{60}	10^{80}	10^{100}	10^{130}	10^{1000}
Lower bound	$37 \cdot 10^{56}$	$28 \cdot 10^{76}$	$22 \cdot 10^{96}$	$17 \cdot 10^{126}$	$21 \cdot 10^{995}$
Upper / Lower	1.97	1.78	1.68	1.53	1.14

Only two possible values

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]

“The Values”

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. $d < 2k \log k$.

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

$$377145549067226075809014239493833600551612641764765068157\text{5}$$

or

$$377145549067226075809014239493833600551612641764765068157\text{6}$$

One value

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

One value

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

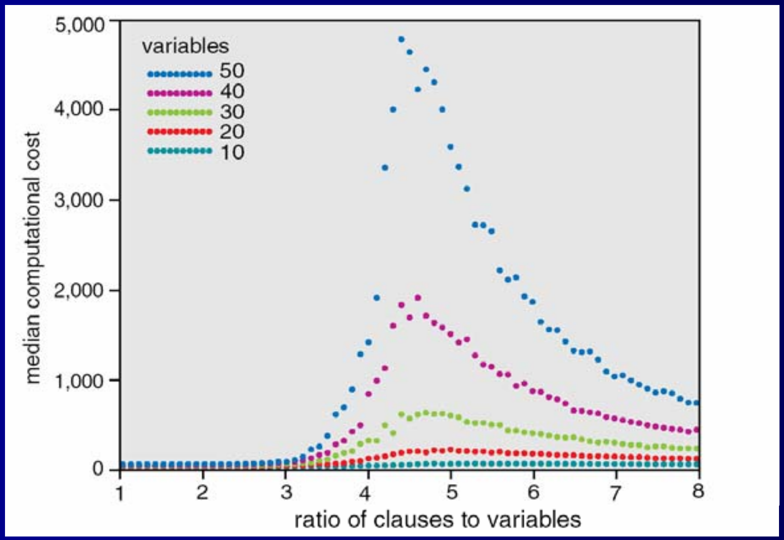
Random k-SAT

- Fix a set of n variables $X = \{x_1, x_2, \ldots, x_n\}$
- Among all $2^k \binom{n}{k}$ possible k-clauses select ***m*** uniformly and independently. Typically $m = rn$.
- Example ($k = 3$) :

$$(\overline{x}_{12} \vee x_5 \vee \overline{x}_9) \wedge (x_{34} \vee \overline{x}_{21} \vee x_5) \wedge \cdots\cdots \wedge (x_{21} \vee x_9 \vee \overline{x}_{13})$$

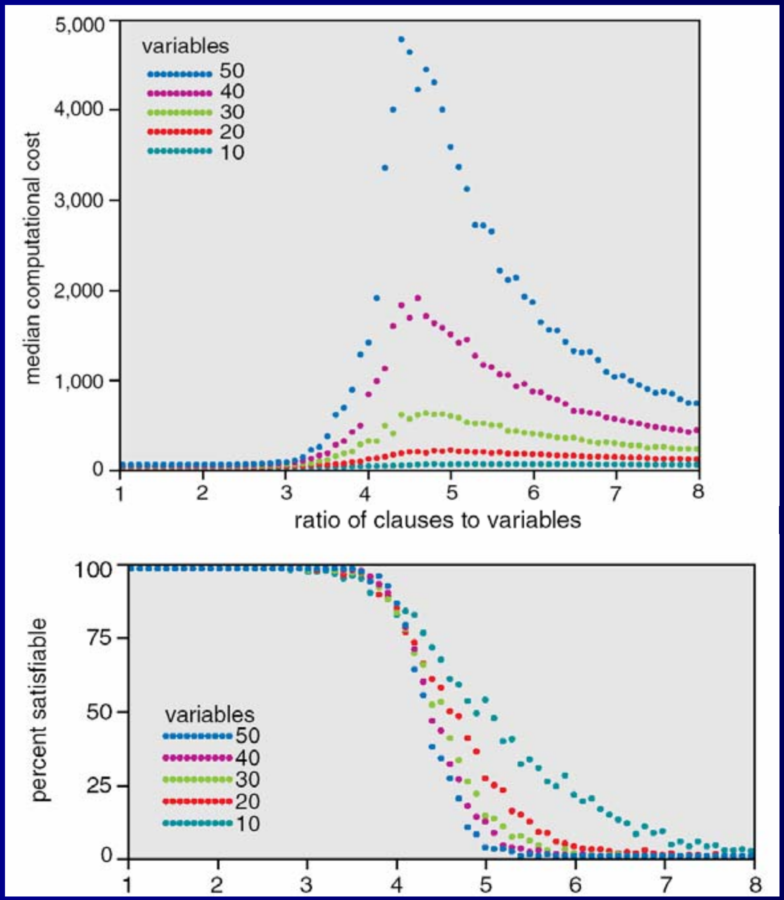
Generating hard 3-SAT instances

[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

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

The satisfiability threshold conjecture

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

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

- 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 it at most

$$2^n\left(1-\frac{1}{2^k}\right)^m = \left(2\left(1-\frac{1}{2^k}\right)^r\right)^n$$

$$\to 0 \quad \text{for } r \geq 2^k \ln 2$$

An asymptotic gap

Since mid-80s, no asymptotic progress over

$$\frac{2^k}{k} < r_k < 2^k \ln 2$$

Getting to within a factor of 2

Theorem: For all $k \geq 3$ and

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

a random k-CNF formula with $m = rn$ 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

k	3	4	5	7	20	21
Upper bound	4.51	10.23	21.33	87.88	726,817	1,453,635
Our lower bound	2.68	7.91	18.79	84.82	726,809	1,453,626
Algorithmic lower bound	3.52	5.54	9.63	33.23	95,263	181,453

The second moment method

For any non-negative r.v. X,

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

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

$$\mathbf{E}[X]^2 = \mathbf{E}[XY]^2 \leq \mathbf{E}[X^2]\mathbf{E}[Y^2] = \mathbf{E}[X^2]\Pr[X > 0] .$$

Ideal for sums

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

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

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

Ideal for sums

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

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

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

Example:

The X_i correspond to the $\binom{n}{q}$ potential q-cliques in $G(n, 1/2)$

Dominant contribution from non-ovelapping 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 $\Longrightarrow$ rapid drop in influence around solutions

General observations

- Method works well when the X_i are like "needles in a haystack"
- Lack of correlations $\Longrightarrow$ rapid drop in influence around solutions
- Algorithms get no "hints"

The second moment method for random k-SAT

- Let X be the # 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

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

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

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

- 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

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

- 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 *σ*

To prove $2^k \ln 2 - k/2 - 1$

- Let $H(\sigma, F)$ be the number of

  satisfied literal occurrences in F under σ

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

$$X(F) = \sum_{\sigma} \mathbf{1}_{\sigma \models F} \, \gamma^{H(\sigma, F)}$$

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

To prove $2^k \ln 2 - k/2 - 1$

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

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

$$
\begin{aligned}
X(F) &= \sum_{\sigma} \mathbf{1}_{\sigma \models F}\, \gamma^{H(\sigma, F)} \\
&= \sum_{\sigma} \prod_{c} \mathbf{1}_{\sigma \models c}\, \gamma^{H(\sigma, c)}
\end{aligned}
$$

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

General functions

- Given any t.a. σ and any k-clause c let

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

be the values of the literals in c under σ.

General functions

- Given any t.a. σ and any k-clause c let

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

be the values of the literals in c under σ.

- We will study random variables of the form

$$X = \sum_{\sigma} \prod_{c} f(\mathbf{v}(\sigma, c))$$

where $f : \{-1, +1\}^k \to \mathbb{R}$ is an arbitrary function

$$X = \sum_{\sigma} \prod_{c} f(\mathbf{v}(\sigma, c))$$

$$f(\mathbf{v}) = 1 \quad \text{for all } \mathbf{v} \quad \Longrightarrow \quad 2^n$$

$$f(\mathbf{v}) = \begin{cases} 0 & \text{if } \mathbf{v} = (-1, -1, \ldots, -1) \\ 1 & \text{otherwise} \end{cases} \quad \Longrightarrow$$

\# of satisfying truth assignments

$$f(\mathbf{v}) = \begin{cases} 0 & \text{if } \mathbf{v} = (-1, -1, \ldots, -1) \text{ or} \\ & \text{if } \mathbf{v} = (+1, +1, \ldots, +1) \\ 1 & \text{otherwise} \end{cases} \quad \Longrightarrow$$

\# of “Not All Equal” truth assignments (NAE)

$$X = \sum_{\sigma} \prod_{c} f(\mathbf{v}(\sigma, c))$$

$$f(\mathbf{v}) = 1 \quad \text{for all } \mathbf{v} \quad \Longrightarrow \quad 2^n$$

$$f(\mathbf{v}) = \begin{cases} 0 & \text{if } \mathbf{v} = (-1, -1, \ldots, -1) \\ 1 & \text{otherwise} \end{cases} \quad \Longrightarrow$$

# of satisfying truth assignments

$$f(\mathbf{v}) = \begin{cases} 0 & \text{if } \mathbf{v} = (-1, -1, \ldots, -1) \text{ or} \\ & \text{if } \mathbf{v} = (+1, +1, \ldots, +1) \\ 1 & \text{otherwise} \end{cases} \quad \Longrightarrow$$

# of satisfying truth assignments whose complement is also satisfying

Overlap parameter = distance

- Overlap parameter is Hamming distance

Overlap parameter = distance

- Overlap parameter is Hamming distance
- For any *f*, if σ, τ agree on $z = n/2$ variables

$$\mathbf{E}\left\langle f(\mathbf{v}(\sigma, c))\, f(\mathbf{v}(\tau, c)) \right\rangle = \mathbf{E}\left[f(\mathbf{v}(\sigma, c)) \right] \mathbf{E}\left[f(\mathbf{v}(\tau, c)) \right]$$

Overlap parameter = distance

- Overlap parameter is Hamming distance
- For any *f*, if σ, τ agree on $z = n/2$ variables

$$\mathbf{E}\left[f(\mathbf{v}(\sigma, c))\, f(\mathbf{v}(\tau, c)) \right] = \mathbf{E}\left[f(\mathbf{v}(\sigma, c)) \right] \mathbf{E}\left[f(\mathbf{v}(\tau, c)) \right]$$

- For any *f* , if σ, τ agree on z variables, let

$$\mathcal{C}_f\,(z/n) \equiv \mathbf{E}\left[f(\mathbf{v}(\sigma, c))\, f(\mathbf{v}(\tau, c)) \right]$$

Contribution according to distance

$$\mathbf{E}[X^2] = \sum_{\sigma,\tau} \prod_{c} \mathbf{E}\left[f(\sigma,c)f(\tau,c)\right]$$

Independence

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

Identically distributed

$$= 2^n \sum_{z=0}^{n} \binom{n}{z} \mathcal{C}_f(z/n)^m$$

Fixing σ

Entropy vs. correlation

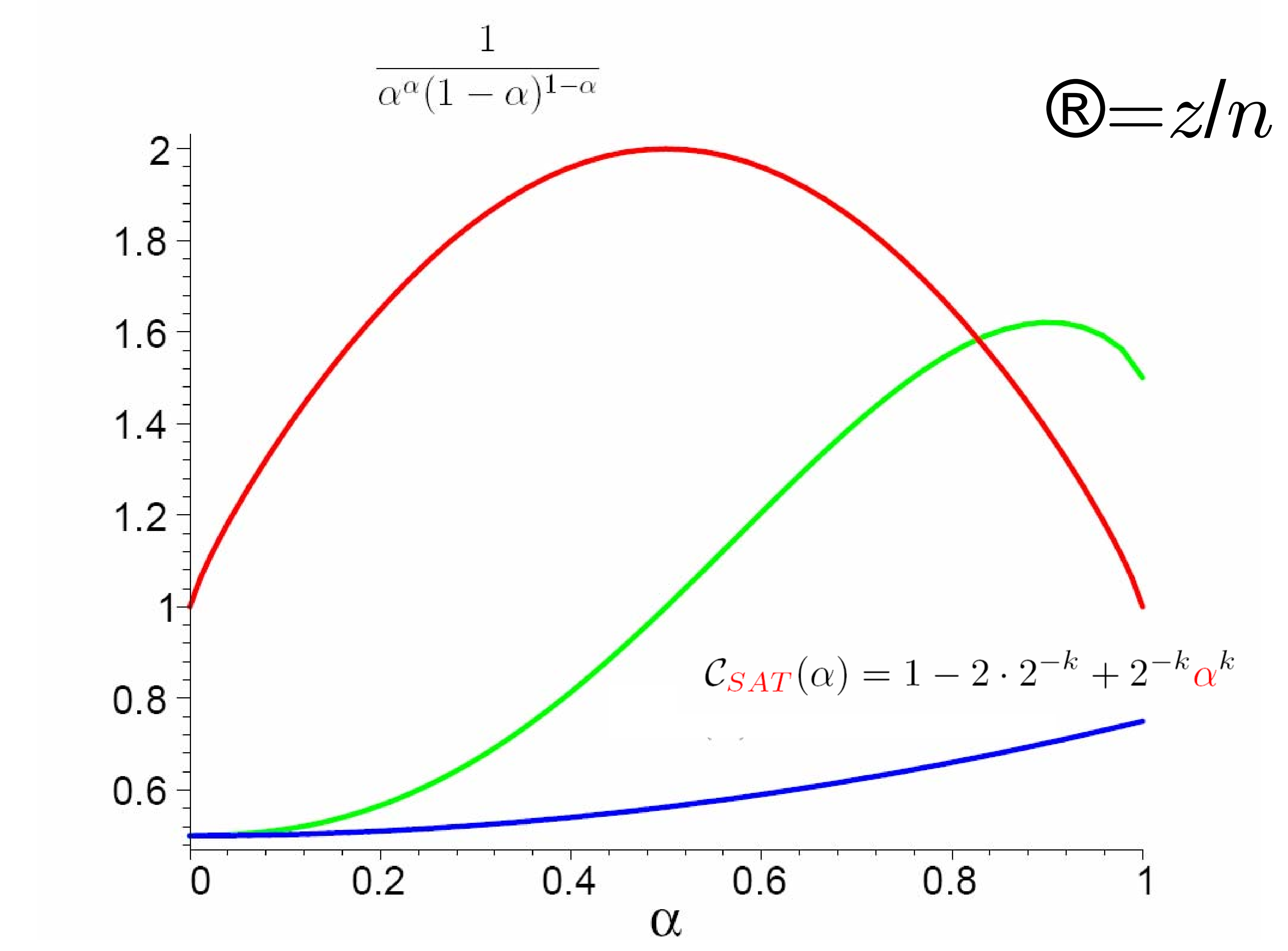
For every function *f* :

$$\mathbf{E}[X^2] = 2^n \sum_{z=0}^{n} \binom{n}{z} \mathcal{C}_f(z/n)^m$$

$$\mathbf{E}[X]^2 = 2^n \sum_{z=0}^{n} \binom{n}{z} \mathcal{C}_f(1/2)^m$$

Contribution according to distance

$$\begin{aligned} \mathbf{E}[X^2] &= \sum_{\sigma,\tau} \prod_{c} \mathbf{E}\left[f(\sigma,c) f(\tau,c)\right] && \textbf{Independence} \\ &= \sum_{\sigma,\tau} \Big(\mathbf{E}\left[f(\sigma,c) f(\tau,c)\right]\Big)^m && \text{Identically distributed} \\ &= 2^n \sum_{z=0}^{n} \binom{n}{z} \mathcal{C}_f(z/n)^m && \text{Fixing } \sigma \\ &= \left(\max_{0 \le \alpha \le 1} \frac{2\,\mathcal{C}_f(\alpha)^r}{\alpha^\alpha (1-\alpha)^{1-\alpha}}\right)^n \times \Theta(1) && \text{Laplace method} \end{aligned}$$



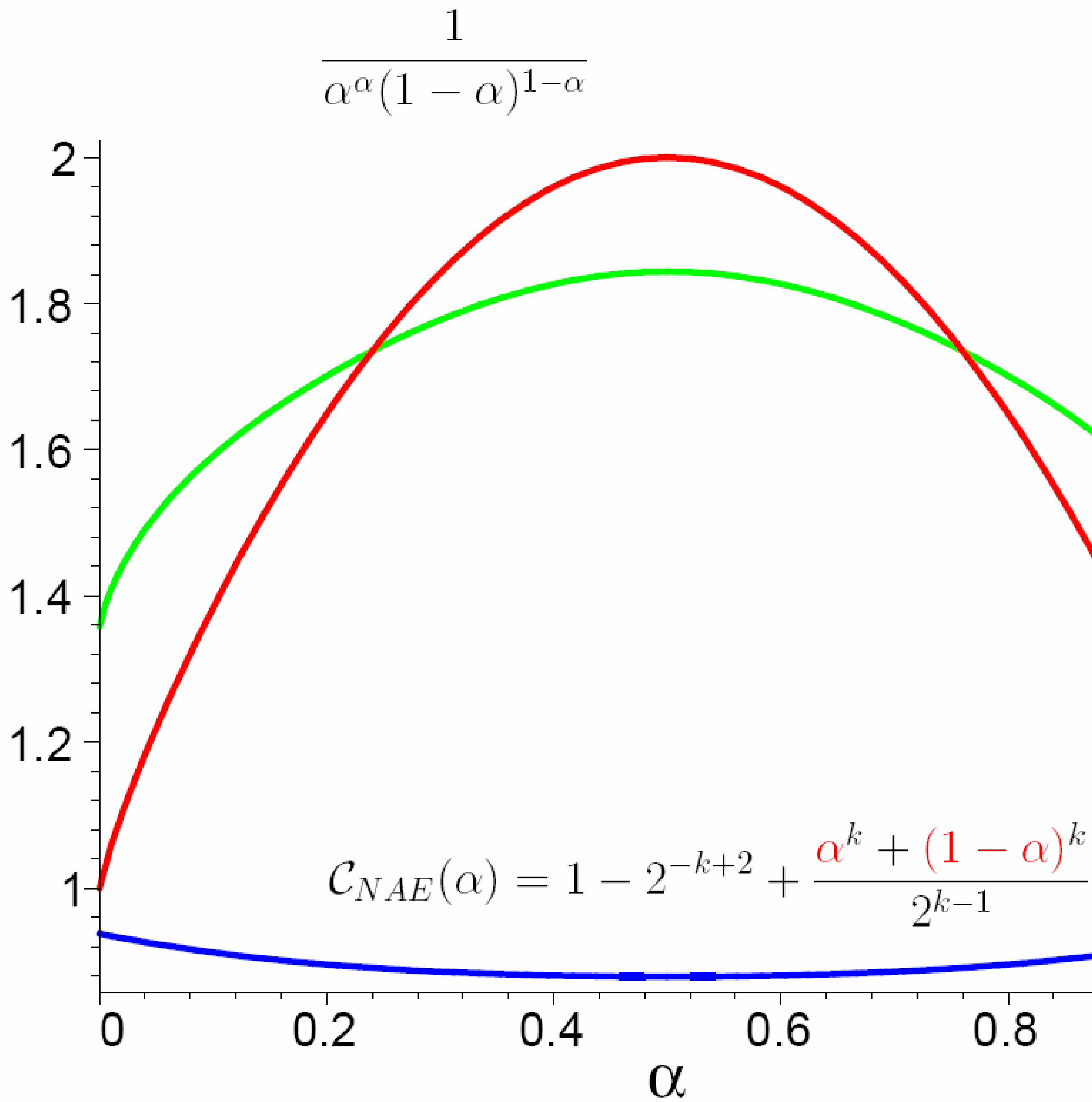
$\frac{1}{\alpha^\alpha(1-\alpha)^{1-\alpha}}$
®=z/n
2
1.8
1.6
1.4
1.2
1
0.8
0.6
0
0.2
0.4
0.6
0.8
1
α
$\mathcal{C}_{SAT}(\alpha) = 1 - 2\cdot 2^{-k} + 2^{-k}\alpha^k$

The importance of being balanced

- An analytic condition:

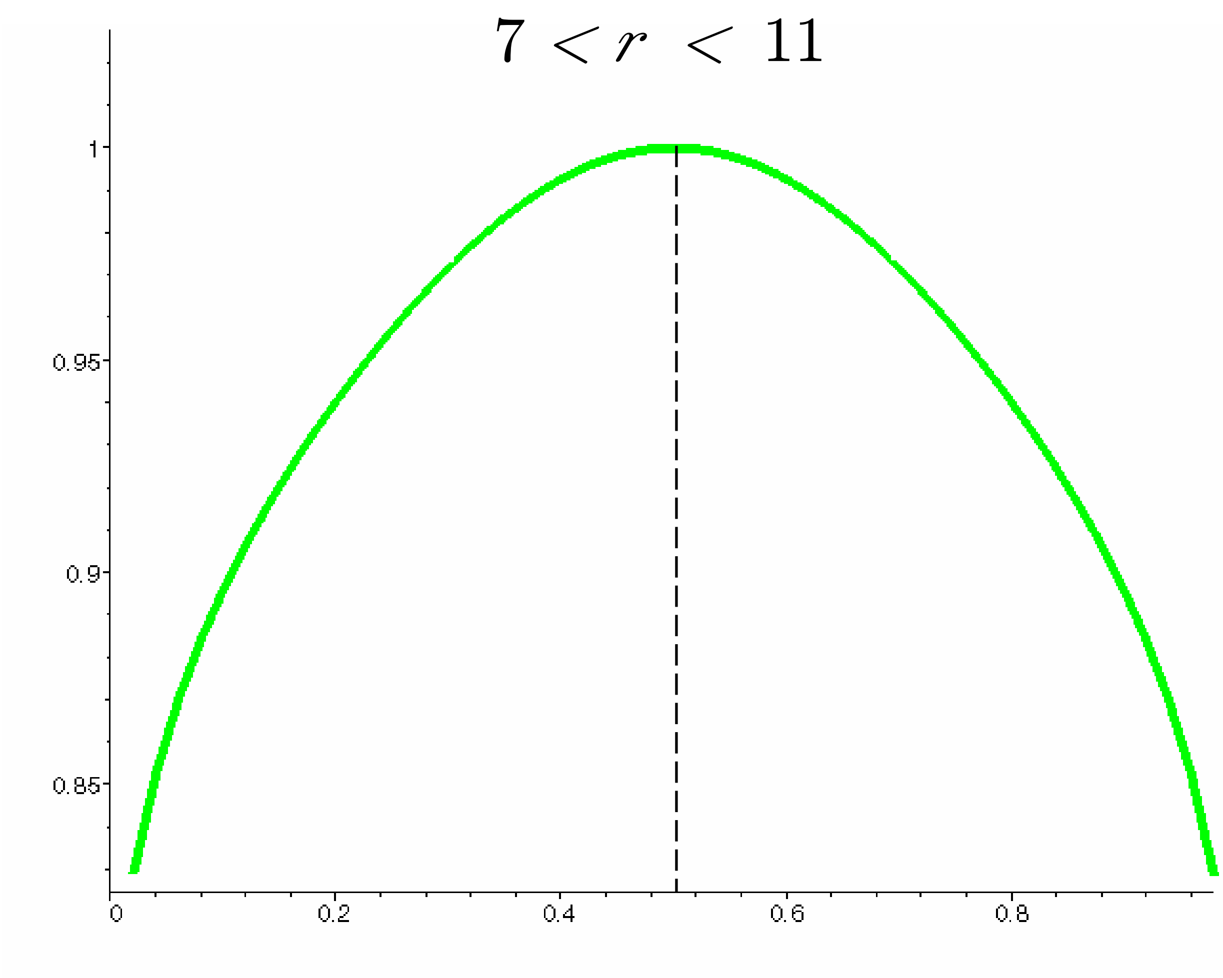
$$\mathcal{C}_f(1/2) \neq 0 \implies \text{the s.m.m. fails}$$

® $= z/n$

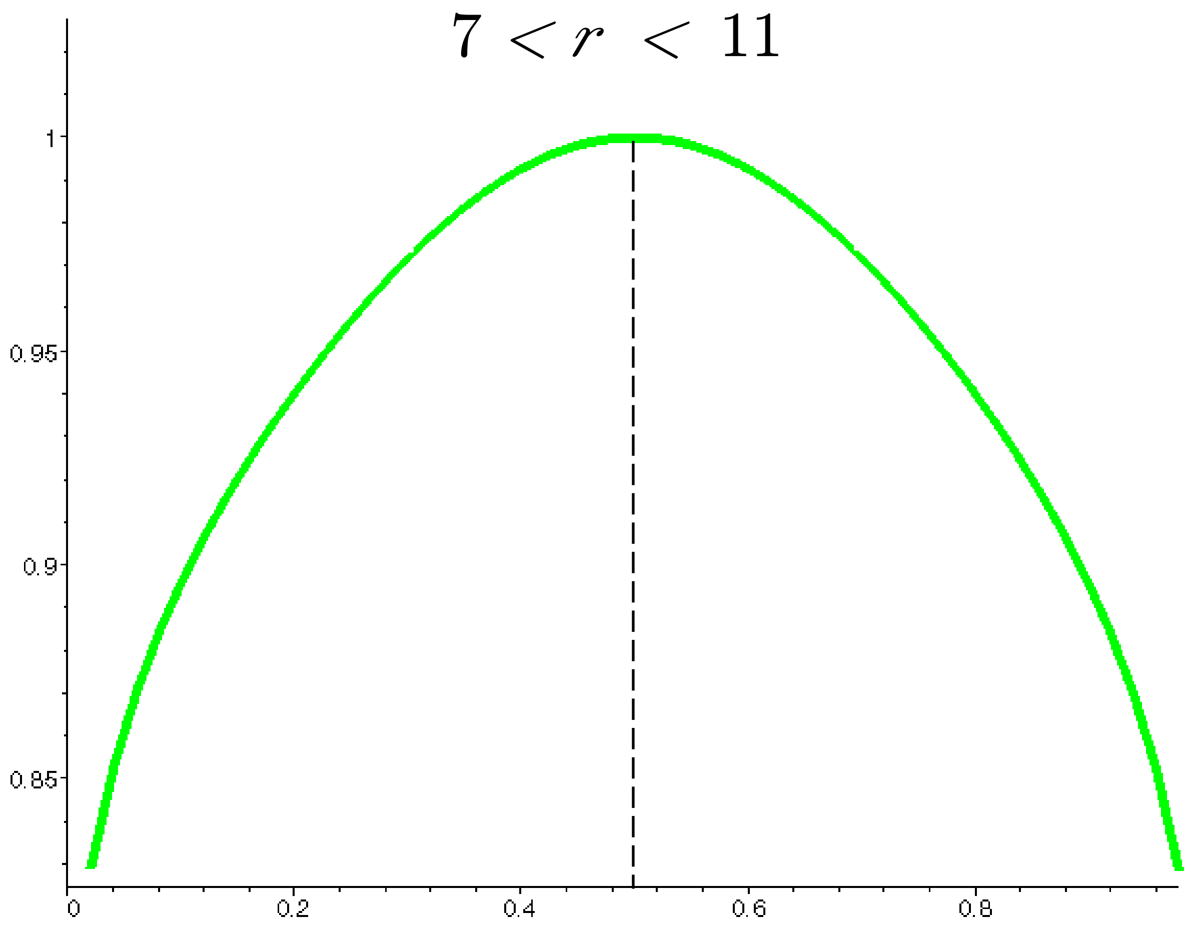


$\frac{1}{\alpha^{\alpha}(1-\alpha)^{1-\alpha}}$
2
1.8
1.6
1.4
1.2
1
$\mathcal{C}_{NAE}(\alpha) = 1 - 2^{-k+2} + \frac{\alpha^k + (1-\alpha)^k}{2^{k-1}}$
0
0.2
0.4
0.6
0.8
1
α

NAE 5-SAT



NAE 5-SAT



The importance of being balanced

- An analytic condition:

$$\mathcal{C}_f(1/2) \neq 0 \implies \text{the s.m.m. fails}$$

The importance of being balanced

- An analytic condition:

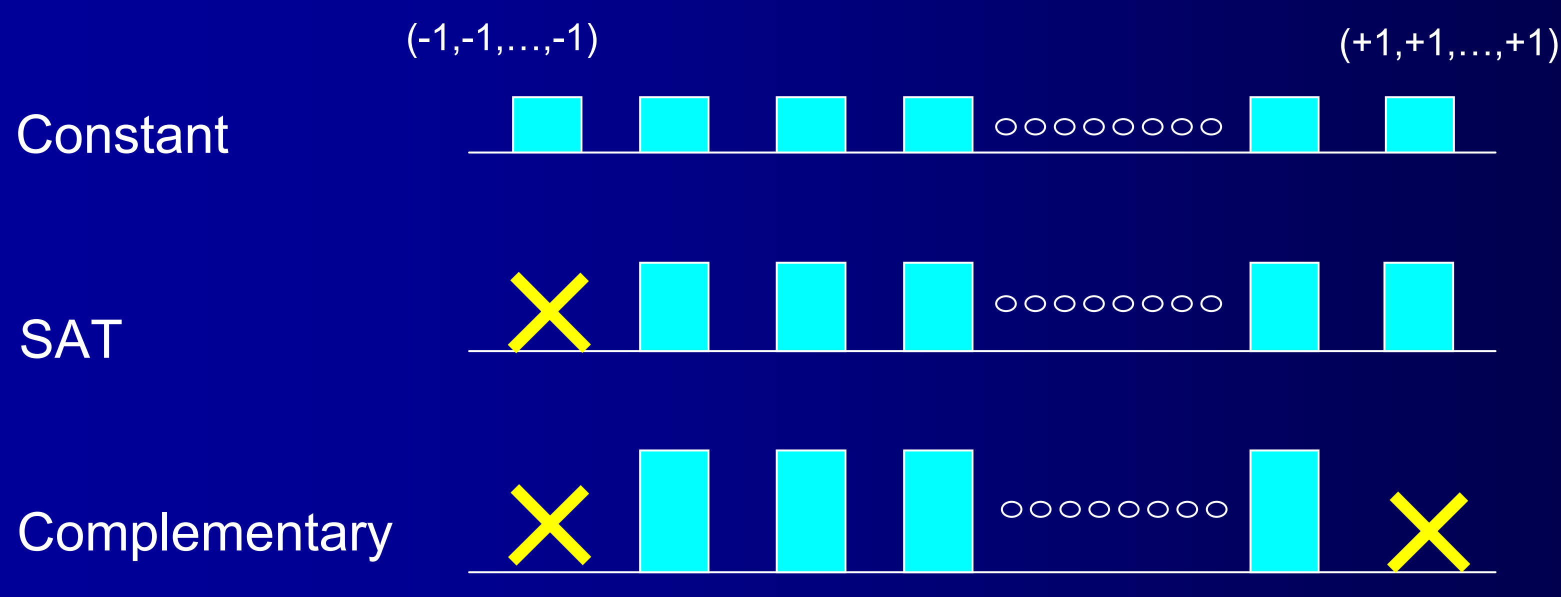
$$\mathcal{C}_f(1=2) \neq 0 \quad \Longrightarrow \quad \text{the s.m.m. fails}$$

- A geometric criterion:

$$\mathcal{C}_f(1=2) = 0 \quad \Longleftrightarrow \quad \sum_{\mathbf{v}\in\{-1;+1\}^k} f(\mathbf{v})\mathbf{v} = 0$$

The importance of being balanced

$$\mathcal{C}_f(1/2) = 0 \iff \sum_{\mathbf{v}\in\{-1;+1\}^k} f(\mathbf{v})\mathbf{v} = 0$$



Balance & Information Theory

- Want to balance vectors in "optimal" way

Balance & Information Theory

- Want to balance vectors in "optimal" way
- Information theory $\Longrightarrow$

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

$$f(-1,-1,\ldots,-1)=0 \quad \text{and} \quad \sum_{\mathbf{v}\in\{-1,+1\}^k} f(\mathbf{v})\mathbf{v}=0$$

Balance & Information Theory

- Want to balance vectors in "optimal" way
- Information theory $\Longrightarrow$

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

$$f(-1,-1,\ldots,-1) = 0 \quad \text{and} \quad \sum_{\mathbf{v} \in \{-1,+1\}^k} f(\mathbf{v})\mathbf{v} = 0$$

- Lagrange multipliers $\Longrightarrow$ the optimal *f* is

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

for the unique γ that satisfies the constraints

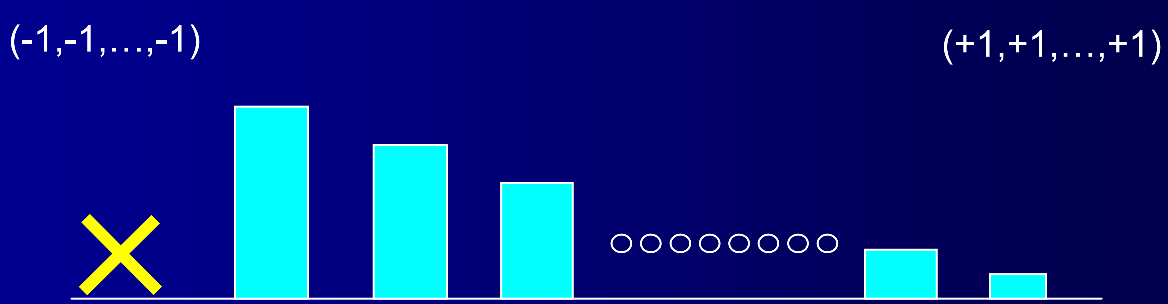
Balance & Information Theory

- Want to balance vectors in "optimal" way
- Information theory $\Longrightarrow$

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

$$f(-1,-1,\ldots,-1) = 0 \quad \text{and} \quad \sum_{\mathbf{v} \in \{-1,+1\}^k} f(\mathbf{v})\mathbf{v} = 0$$

Heroic



Random graph coloring

Threshold formulation

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

$$c \le k \log k - \log k - 1$$

and w□h□p□ non-k-colorable if

$$c \ge k \log k - \frac{1}{2} \log k \quad .$$

Main points

- Non-*k*-colorability:

  Proof. *The probability that there exists any k-coloring is at most* $k^n \left(1 - \frac{1}{k}\right)^{cn} \to 0$

- *k*-colorability:

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

Setup

- Let X_σ be the indicator that the balanced k-partition σ is a proper k-coloring.

- We will prove that if $X = \sum_{\sigma} X_{\sigma}$ then for all $c \leq k \log k - \log k - 1$ there is a constant $D = D(k) > 0$ such that

$$\mathbf{E}[X^2] < D\,\mathbf{E}[X]^2$$

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

Setup

- $\mathbf{E}[X^2] =$ sum over all σ, τ of $\mathbf{E}[X_\sigma X_\tau]$.

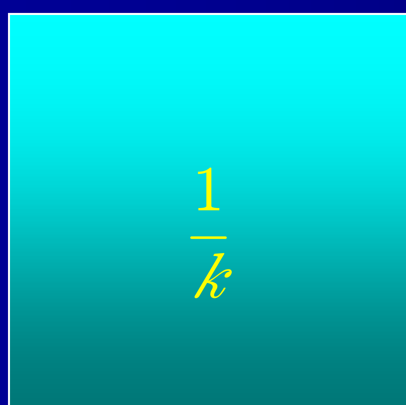
- For any pair of balanced k-partitions σ, τ let $a_{ij}n$ be the # of vertices having color i in σ and color j in τ.

$$\Pr[\sigma \text{ and } \tau \text{ are proper}] = \left(1 - \frac{2}{k} + \sum_{ij} a_{ij}^2\right)^{cn}$$

Examples

Balance $\Longrightarrow$ A is doubly-stochastic.

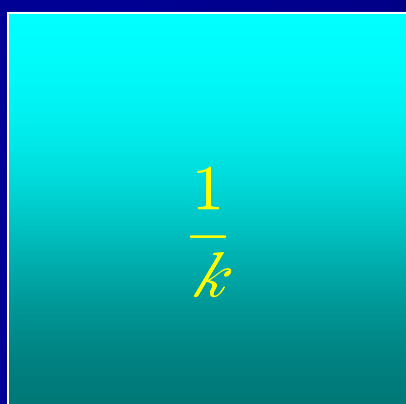
When σ, τ are uncorrelated, A is the flat $1/k$ matrix



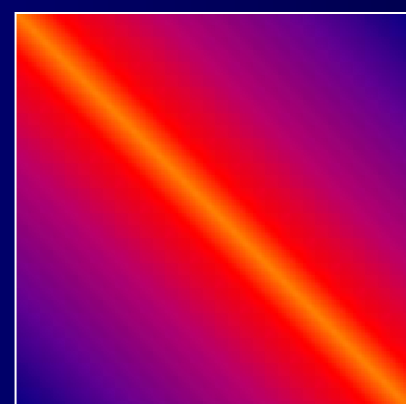
Examples

Balance $\Longrightarrow$ A is doubly-stochastic.

When σ, τ are uncorrelated, A is the flat $1/k$ matrix



As σ, τ align, A tends to the identity matrix I



A matrix-valued overlap

So, $$\mathbf{E}[X^2] = \sum_{A \in B_k} \binom{n}{An} \left(1 - \frac{2}{k} + \frac{1}{k^2} \sum d_{ij}^2 \right)^{cn}$$

A matrix-valued overlap

So, $\mathbf{E}[X^2] = \displaystyle\sum_{A \in B_k} \binom{n}{An} \left(1 - \frac{2}{k} + \frac{1}{k^2} \sum a_{ij}^2 \right)^{cn}$

which is controlled by the maximizer of

$$-\sum a_{ij} \log a_{ij} + c \log \left(1 - \frac{2}{k} + \frac{1}{k^2} \sum a_{ij}^2 \right)$$

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

A matrix-valued overlap

So, $$\mathbf{E}[X^2] = \sum_{A \in B_k} \binom{n}{An} \left(1 - \frac{2}{k} + \frac{1}{k^2} \sum d_{ij}^2 \right)^{cn}$$

which is controlled by the maximizer of

$$-\sum a_{ij} \log a_{ij} + c \qquad \frac{1}{k^2} \sum d_{ij}^2$$

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

$$- \sum_{i,j} a_{ij} \log a_{ij} + c \cdot \frac{1}{k^2} \sum_{i,j} d_{ij}^2$$

- Entropy decreases away from the flat $1/k$ matrix

$$-\sum_{i,j} a_{ij} \log a_{ij} + c \cdot \frac{1}{k^2} \sum_{i,j} d_{ij}^2$$

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

$$-\sum_{i,j} a_{ij} \log a_{ij} + c \cdot \frac{1}{k^2} \sum_{i,j} d_{ij}^2$$

- 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,….

$$-\sum_{i,j} a_{ij} \log a_{ij} + c \cdot \frac{1}{k^2} \sum_{i,j} d_{ij}^2$$

- 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

$$-\sum_{i,j} a_{ij} \log a_{ij} + c \cdot \frac{1}{k^2} \sum_{i,j} d_{ij}^2$$

- 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 overview

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 ρ_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 ρ_i.
3. Find max-entropy, $f(\rho_i)$, of each row given ρ_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 ρ_i.
3. Find max-entropy, $f(\rho_i)$, of each row given ρ_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 L_2 norm of each row ρ_i.
3. Find max-entropy, $f(\rho_i)$, of each row given ρ_i.
4. Prove that $f''' > 0$.
5. Use (4) to determine the optimal distribution of the ρ_i given their total ρ.

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 ρ_i.
3. Find max-entropy, $f(\rho_i)$, of each row given ρ_i.
4. Prove that $f''' > 0$.
5. Use (4) to determine the optimal distribution of the ρ_i given their total ρ.
6. Optimize over ρ.

Random regular graphs

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

$$-\sum_{i=1}^{k} a_i \log a_i$$

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

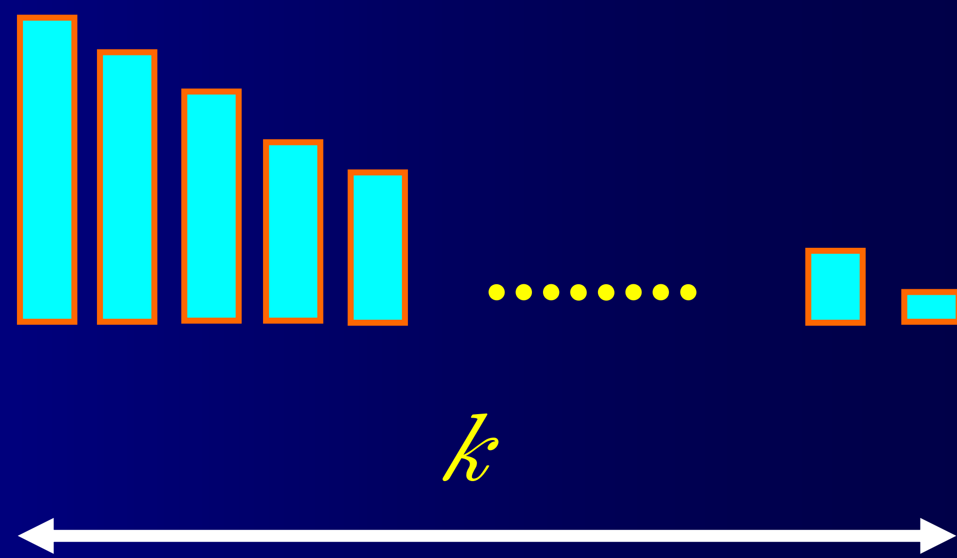
$$-\sum_{i=1}^{k} a_i \log a_i$$

subject to

$$\sum_{i=1}^{k} a_i = 1$$

$$\sum_{i=1}^{k} a_i^2 = \rho$$

for some $1/k < \rho < 1$



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 $x > y$

A vector analogue (optimizing a single row)

Maximize

$$-\sum_{i=1}^{k} a_i \log a_i$$

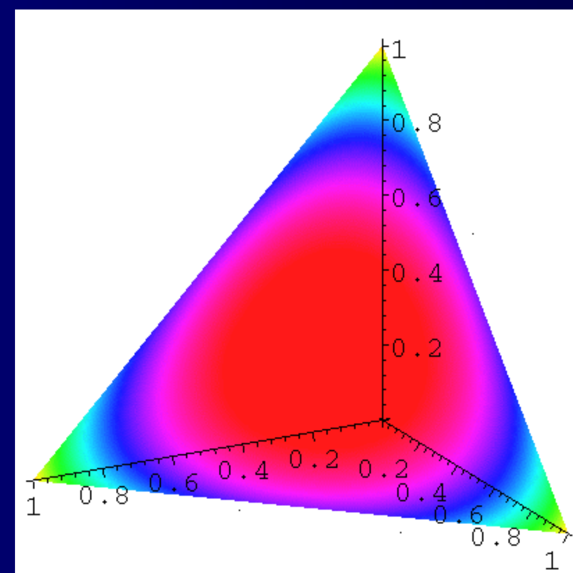
subject to

$$\sum_{i=1}^{k} a_i = 1$$

$$\sum_{i=1}^{k} a_i^2 = \rho$$

for some $1/k < \rho < 1$

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



A vector analogue
(optimizing a single row)

Maximize

$$-\sum_{i=1}^{k} a_i \log a_i$$

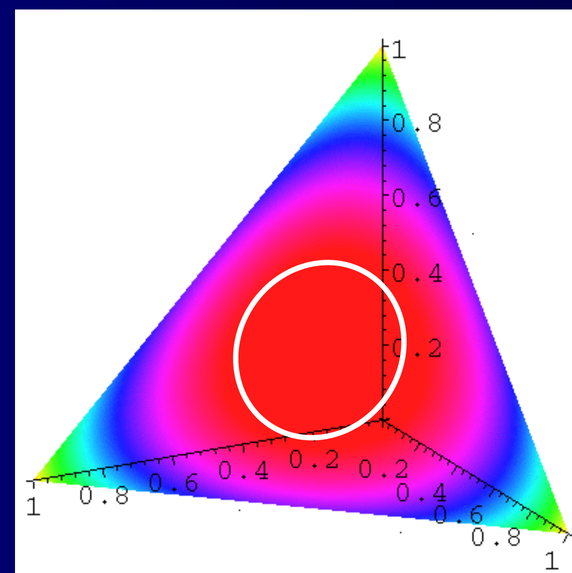
subject to

$$\sum_{i=1}^{k} a_i = 1$$

$$\sum_{i=1}^{k} a_i^2 = \rho$$

for some $1/k < \rho < 1$

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



A vector analogue (optimizing a single row)

Maximize

$$-\sum_{i=1}^{k} a_i \log a_i$$

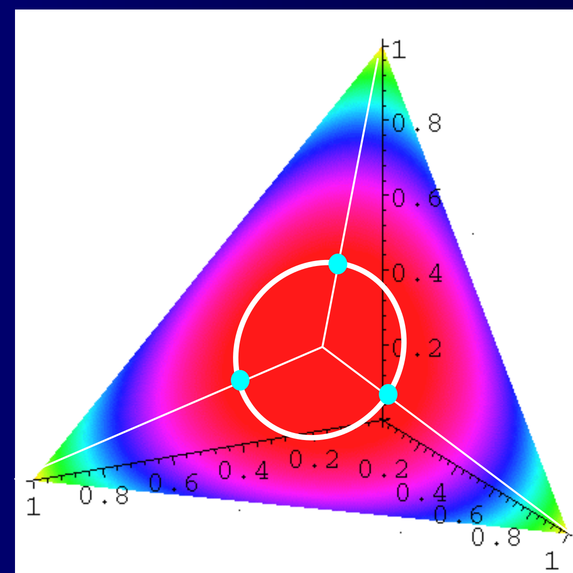
subject to

$$\sum_{i=1}^{k} a_i = 1$$

$$\sum_{i=1}^{k} a_i^2 = \rho$$

for some $1/k < \rho < 1$

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



A vector analogue
(optimizing a single row)

Maximize

$$-\sum_{i=1}^{k} a_i \log a_i$$

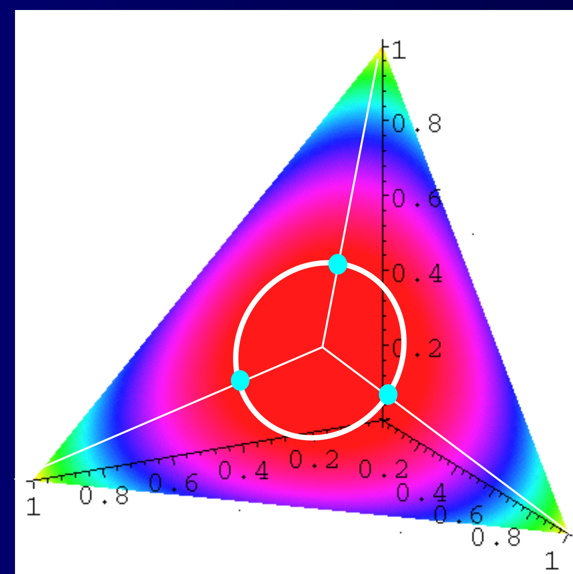
subject to

$$\sum_{i=1}^{k} a_i = 1$$

$$\sum_{i=1}^{k} a_i^2 = \rho$$

for some $1/k < \rho < 1$

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



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