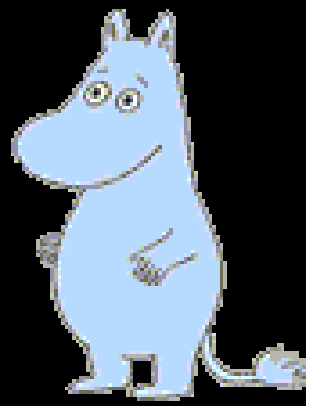
Adaptive annealing: a near-optimal connection between sampling and counting

Daniel Štefankovič
(University of Rochester)

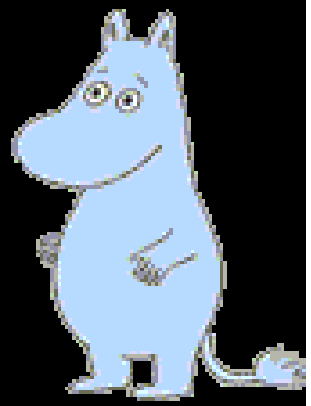
Santosh Vempala
Eric Vigoda
(Georgia Tech)



Adaptive ap[illegible]r-optimal

If you want to count using MCMC then statistical physics is useful.

[illegible]goda
(Georgia Tech)



Outline

1. **Counting problems**

2. Basic tools: Chernoff, Chebyshev

3. Dealing with large quantities
   (the product method)

4. Statistical physics

5. Cooling schedules (our work)

6. More…

Counting

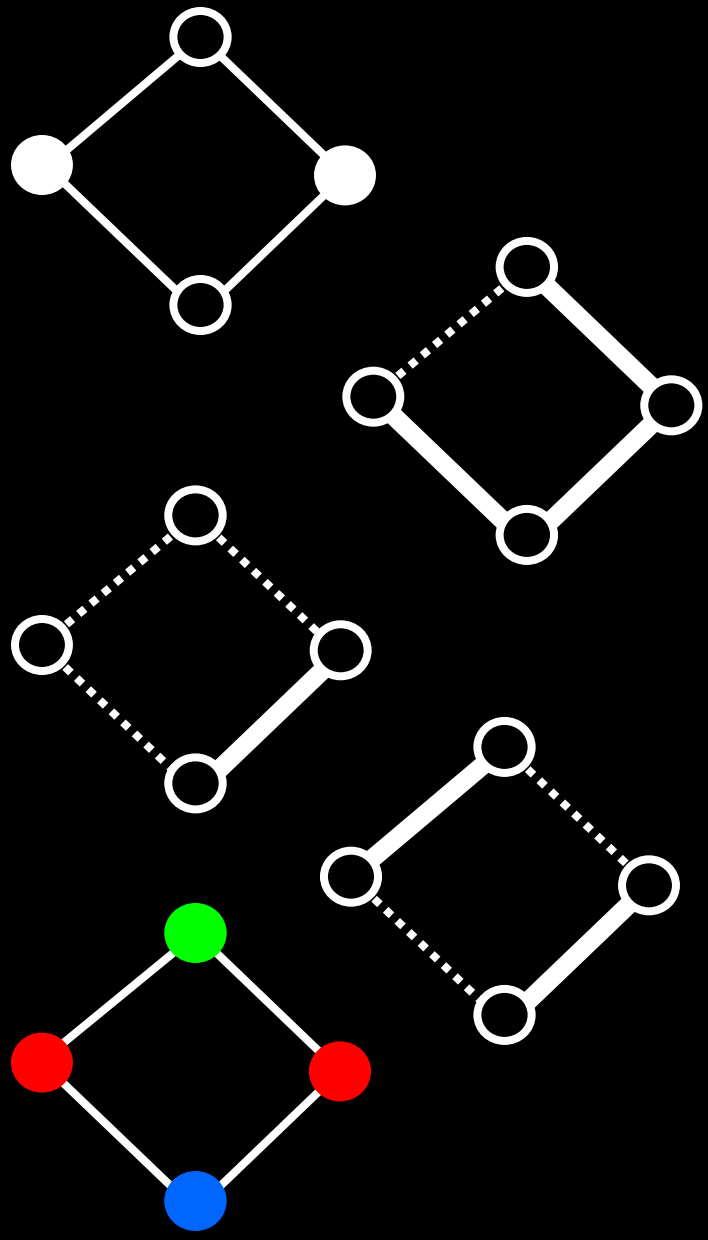
independent sets

spanning trees

matchings

perfect matchings

k-colorings



Counting

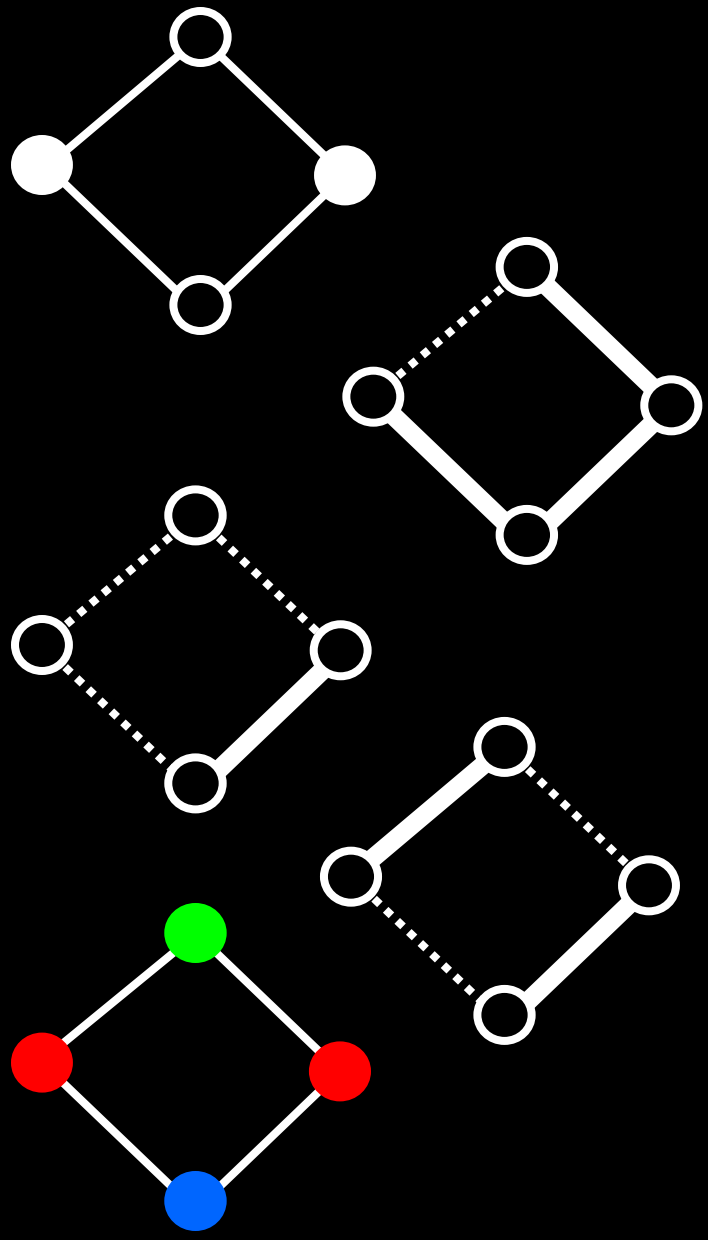
independent sets

spanning trees

matchings

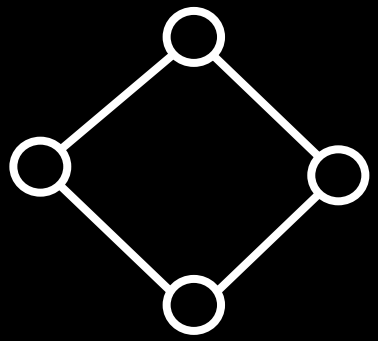
perfect matchings

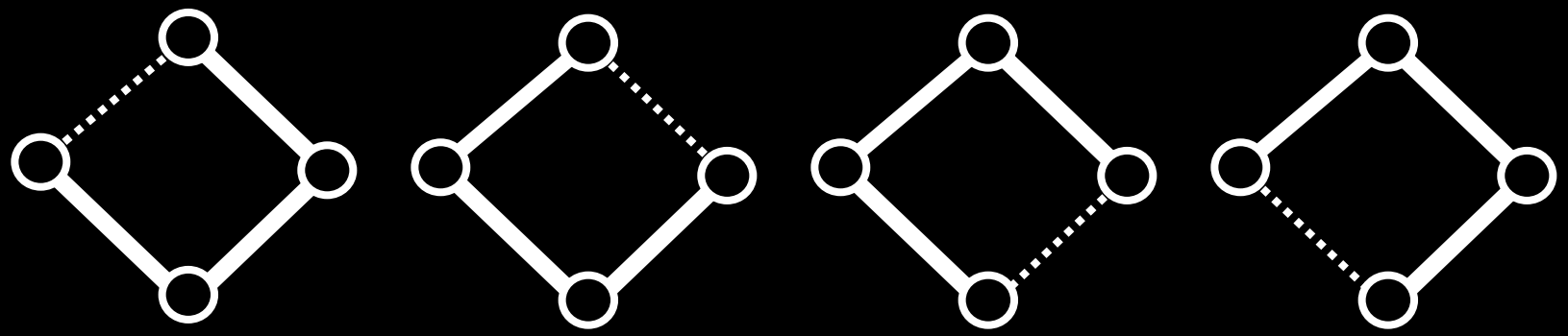
k-colorings



Compute the number of

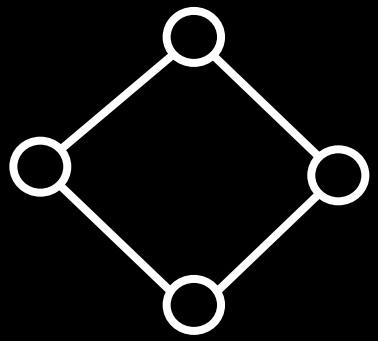
spanning trees





Compute the number of

spanning trees



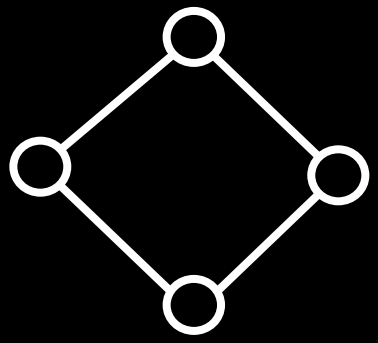
Kirchhoff's Matrix Tree Theorem: $\det(D - A)_{vv}$

$$\begin{pmatrix} 2 & 0 & 0 & 0 \\ 0 & 2 & 0 & 0 \\ 0 & 0 & 2 & 0 \\ 0 & 0 & 0 & 2 \end{pmatrix} - \begin{pmatrix} 0 & 1 & 0 & 1 \\ 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 \\ 1 & 0 & 1 & 0 \end{pmatrix}$$

D A

$$\det \begin{pmatrix} 2 & -1 & 0 \\ -1 & 2 & -1 \\ 0 & -1 & 2 \end{pmatrix}$$

Compute the number of

spanning trees 

G → polynomial-time algorithm

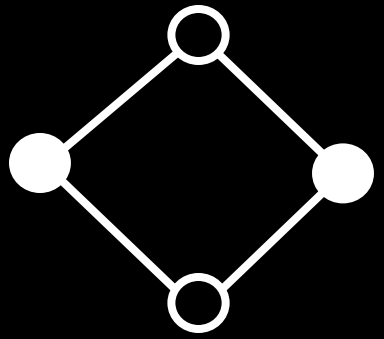
↓

number of spanning trees of G

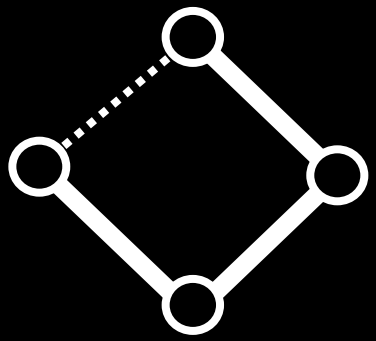
Counting

?

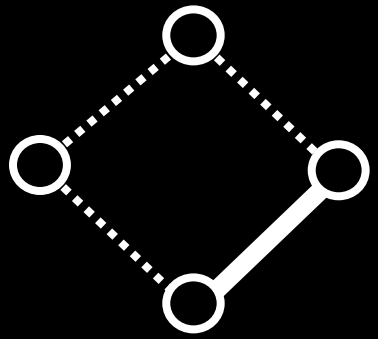
independent sets



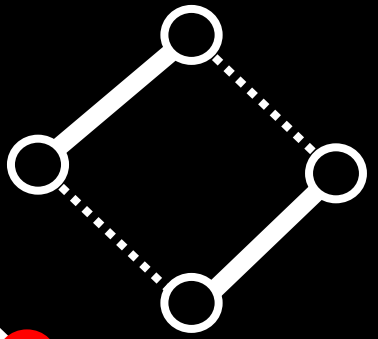
spanning trees



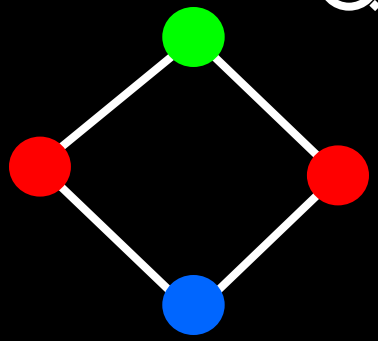
matchings



perfect matchings

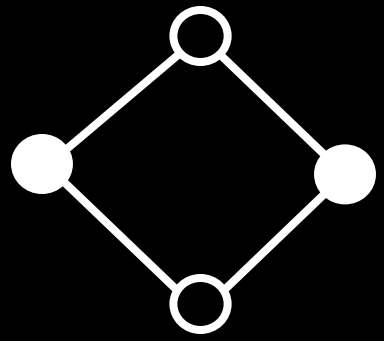


k-colorings

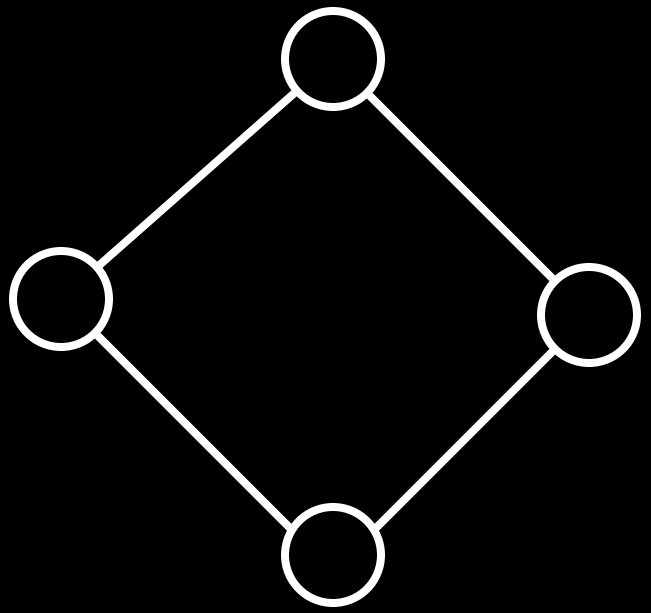


Compute the number of

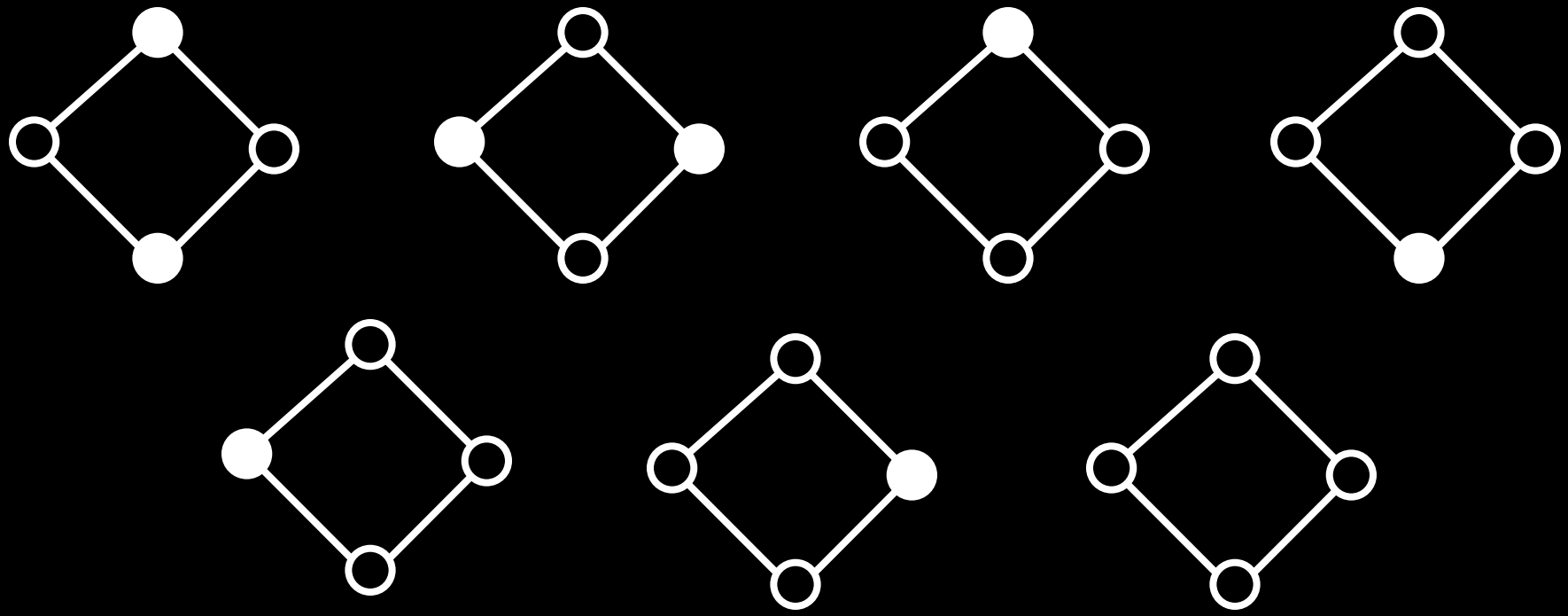
independent sets
(hard-core gas model)



independent set of a graph = subset S of vertices, no two in S are neighbors

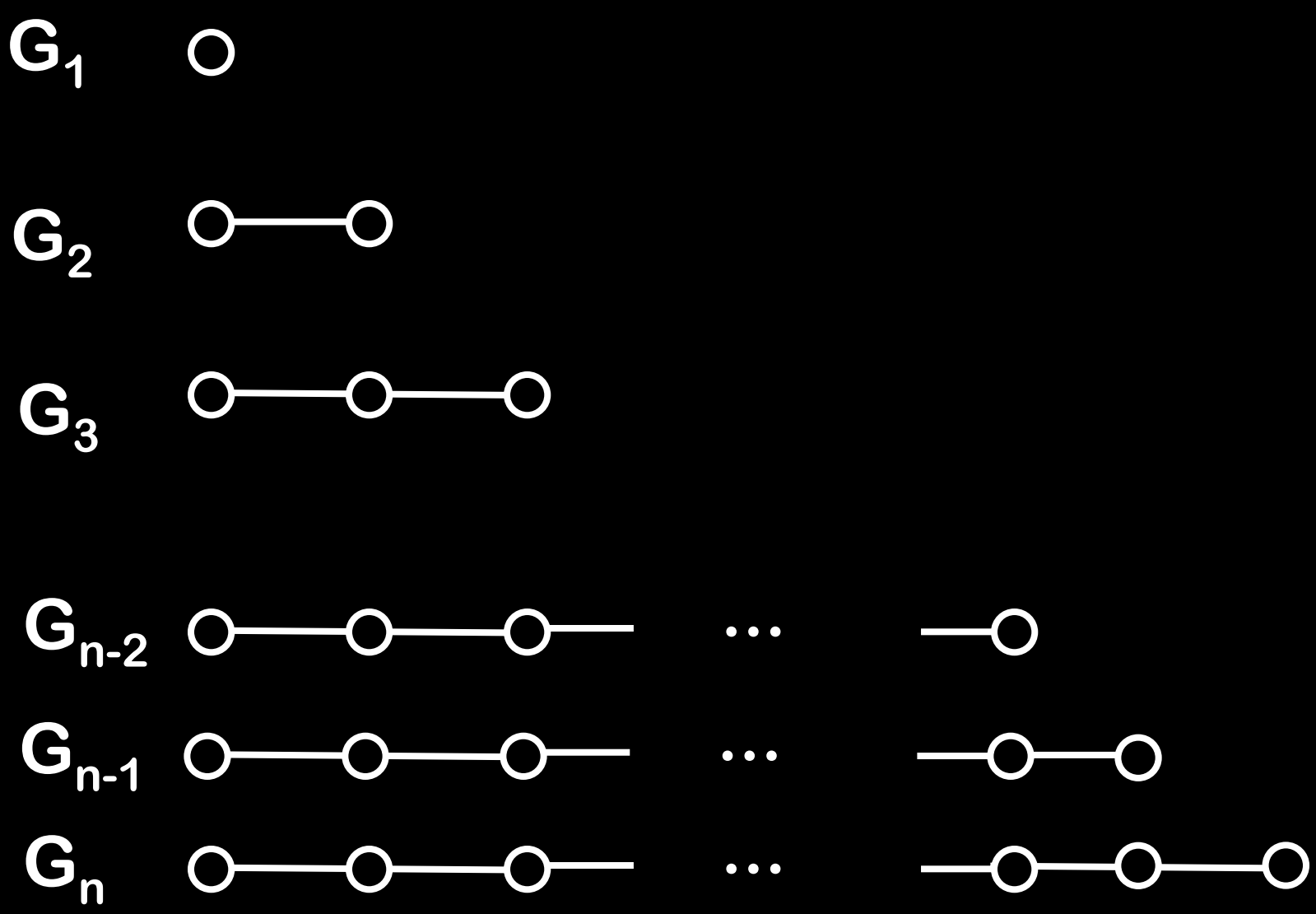


independent sets = 7

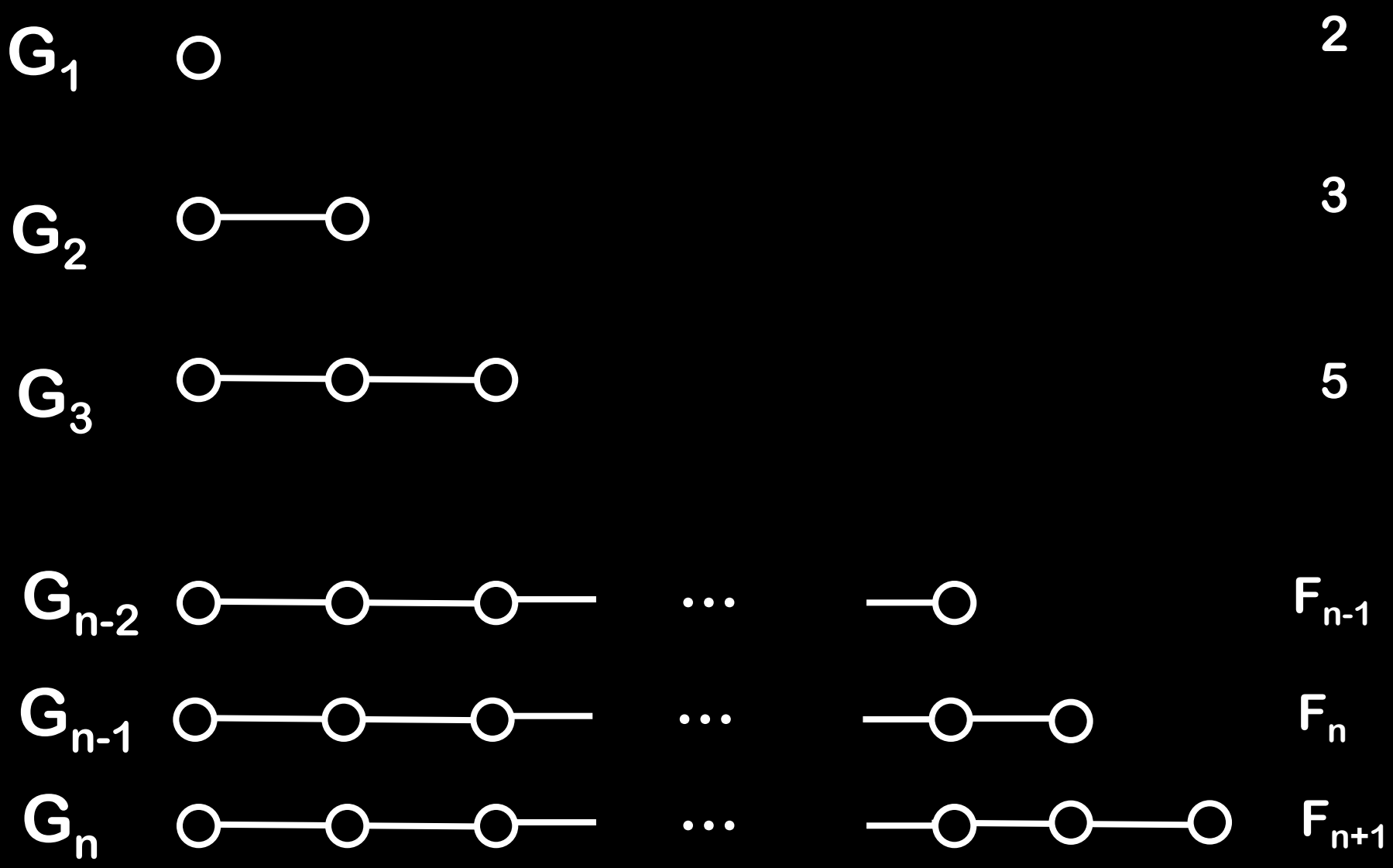


independent set = subset S of vertices
no two in S are neighbors

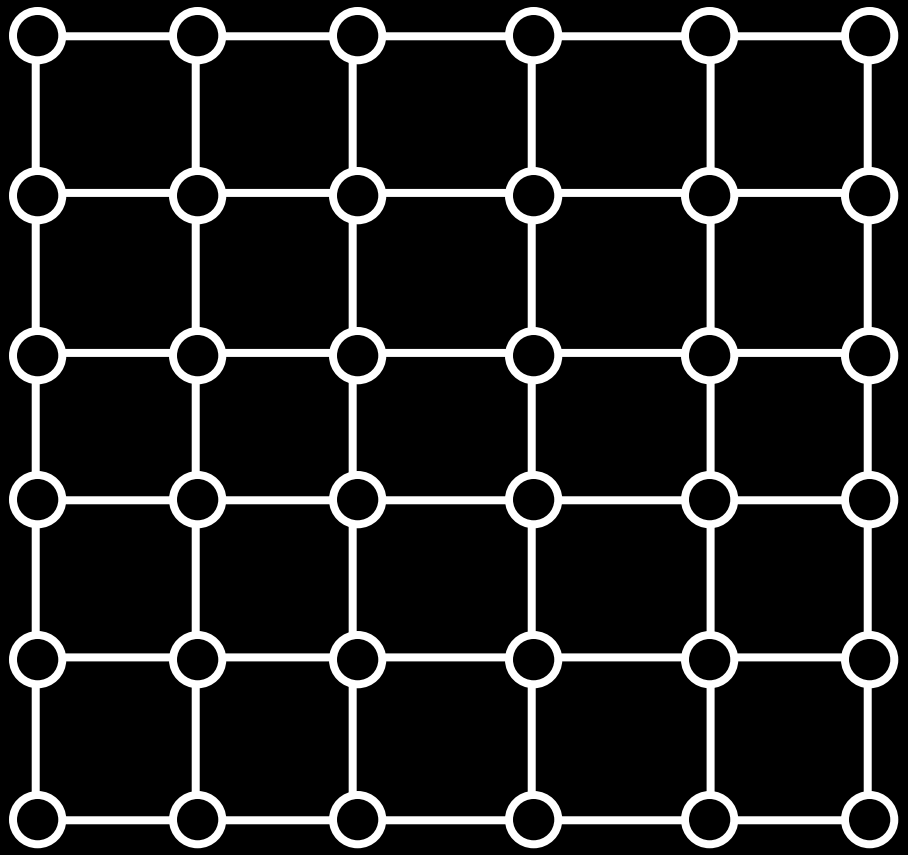
independent sets =



independent sets =



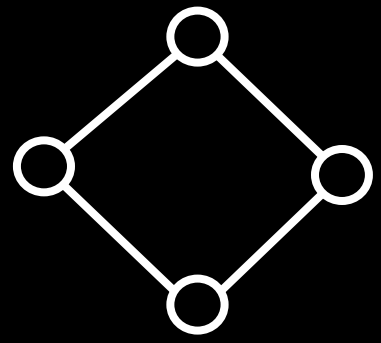
independent sets = 5598861



independent set = subset S of vertices
no two in S are neighbors

Compute the number of

independent sets

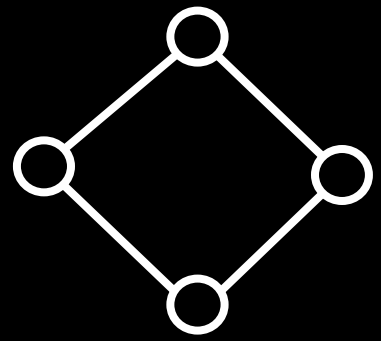


G → polynomial-time algorithm → ?

number of independent sets of G

Compute the number of

independent sets



(unlikely)

G →

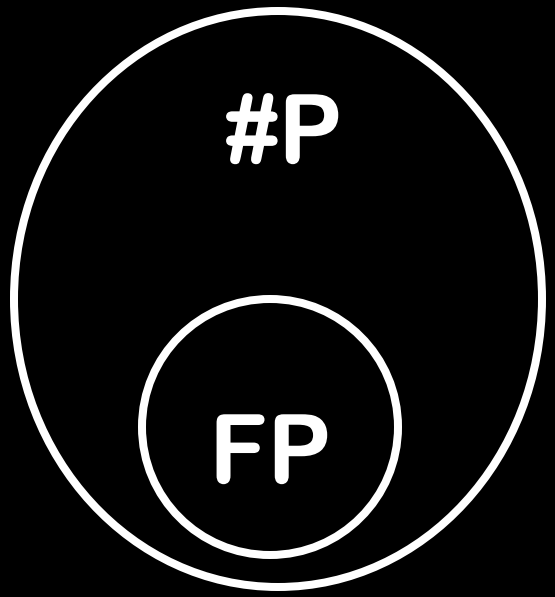
polynomial-time algorithm

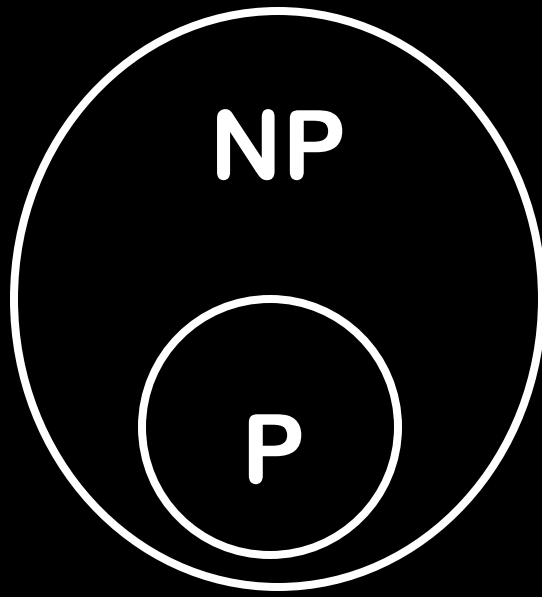
!

↓

number of independent sets of G

graph G $\mapsto$ # independent sets in G





#P-complete

#P-complete even for 3-regular graphs

(Dyer, Greenhill, 1997)

graph $G \mapsto$ # independent sets in G

?

approximation

randomization

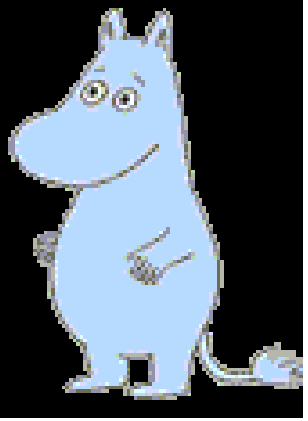
graph G $\mapsto$ # independent sets in G

?

approximation

randomization

which is
more important?



graph G $\mapsto$ # independent sets in G

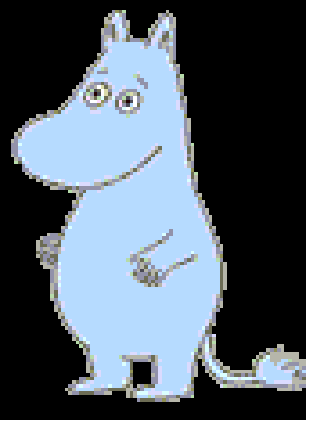
My world-view:

(true) randomness is important conceptually but NOT computationally (i.e., I believe P=BPP).

approximation makes problems easier (i.e., I believe #P≠BPP)

important?

randomization



We would like to know Q

Goal: random variable Y such that

$$P((1-\varepsilon)Q \leq Y \leq (1+\varepsilon)Q) \geq 1-\delta$$

"Y gives $(1\pm\varepsilon)$-estimate"

We would like to know Q

Goal: random variable Y such that

$$P((1-\varepsilon)Q \leq Y \leq (1+\varepsilon)Q) \geq 1-\delta$$

(fully polynomial randomized approximation scheme):

FPRAS:

$G,\varepsilon,\delta \rightarrow$ polynomial-time algorithm $\rightarrow Y$

Outline

1. Counting problems

2. Basic tools: Chernoff, Chebyshev

3. Dealing with large quantities
   (the product method)

4. Statistical physics

5. Cooling schedules (our work)

6. More...

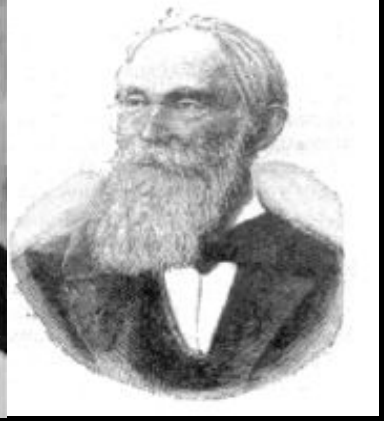
We would like to know Q

1. Get an unbiased estimator X, i. e.,

$$E[X] = Q$$

2. "Boost the quality" of X:

$$Y = \frac{X_1 + X_2 + \ldots + X_n}{n}$$



The Bienaymé-Chebyshev inequality

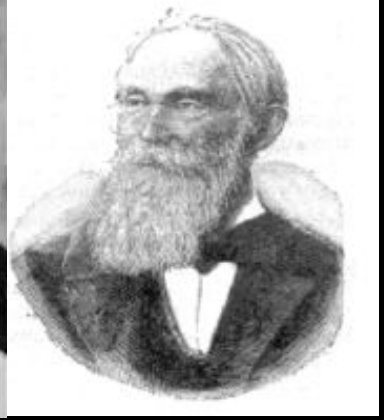
We would like to know Q

Goal: random variable Y such that

$P((1-\varepsilon)Q \le Y \le (1+\varepsilon)Q) \ge 1-\delta$

"Y gives $(1\pm\varepsilon)$-estimate"

$$P(\text{ Y gives } (1\pm\varepsilon)\text{-estimate }) \ge 1 - \frac{V[Y]}{E[Y]^2}\,\frac{1}{\varepsilon^2}$$



The Bienaymé-Chebyshev inequality

We would like to know Q

Goal: random variable Y such that

$$P((1-\varepsilon)Q \le Y \le (1+\varepsilon)Q) \ge 1-\delta$$

"Y gives $(1\pm\varepsilon)$-estimate"

$$P(\text{ Y gives } (1\pm\varepsilon)\text{-estimate }) \ge 1 - \frac{V[Y]}{E[Y]^2}\,\frac{1}{\varepsilon^2}$$

squared coefficient of variation SCV

$$Y = \frac{X_1 + X_2 + \ldots + X_n}{n} \Rightarrow \frac{V[Y]}{E[Y]^2} = \frac{1}{n}\,\frac{V[X]}{E[X]^2}$$

The Bienaymé-Chebyshev inequality

Let $X_1,\ldots,X_n,X$ be independent, identically distributed random variables, $Q=E[X]$. Let

$$Y = \frac{X_1 + X_2 + \ldots + X_n}{n}$$

Then

$$P(\text{ Y gives } (1\pm\varepsilon)\text{-estimate of Q }) \geq 1 - \frac{V[X]}{n\,E[X]^2}\,\frac{1}{\varepsilon^2}$$

Chernoff's bound

Let $X_1,\ldots,X_n,X$ be independent, identically distributed random variables, $0 \le X \le 1$, $Q=E[X]$. Let

$$Y = \frac{X_1 + X_2 + \ldots + X_n}{n}$$

Then

$$P(\text{ } Y \text{ gives } (1\pm\varepsilon)\text{-estimate of } Q \text{ })$$

$$\ge 1 - e^{-\varepsilon^2 \cdot n \cdot E[X] / 3}$$

The Bienaymé-Chebyshev inequality

Let $X_1,\ldots,X_n,X$ be independent, identically distributed random variables, $Q=E[X]$. Let

$$Y = \frac{X_1 + X_2 + \ldots + X_n}{n}$$

Then

$$P(\text{ Y gives } (1\pm\varepsilon)\text{-estimate of Q }) \geq 1 - \frac{V[X]}{n\,E[X]^2}\,\frac{1}{\varepsilon^2}$$

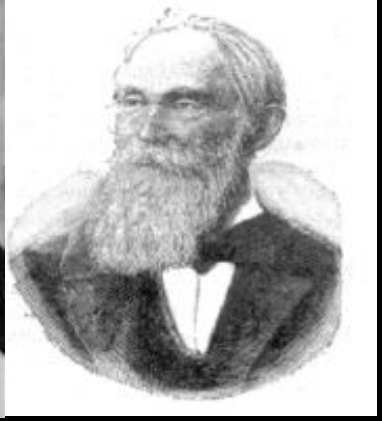
Chernoff's bound

Let $X_1,\ldots,X_n,X$ be independent, identically distributed random variables, $0 \leq X \leq 1$, $Q=E[X]$. Let

$$Y = \frac{X_1 + X_2 + \ldots + X_n}{n}$$

Then

$$P(\text{ Y gives } (1\pm\varepsilon)\text{-estimate of Q }) \geq 1 - e^{-\varepsilon^2 \,.\, n \,.\, E[X] / 3}$$

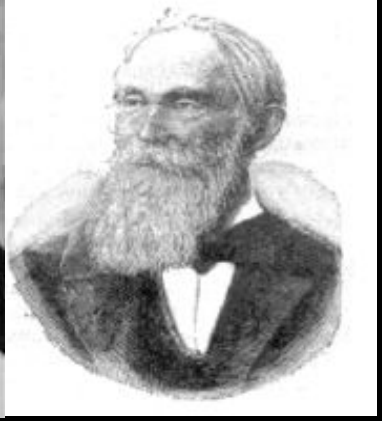


$$n = \frac{V[X]}{E[X]^2} \frac{1}{\varepsilon^2} \frac{1}{\delta}$$

Number of samples to achieve precision ε with confidence δ.

$$n = \frac{1}{E[X]} \frac{3}{\varepsilon^2} \ln(1/\delta)$$

$0 \leq X \leq 1$



$$n = \frac{V[X]}{E[X]^2} \frac{1}{\varepsilon^2} \frac{1}{\delta}$$

BAD

Number of samples to achieve precision ε with confidence δ.

$$n = \frac{1}{E[X]} \frac{3}{\varepsilon^2} \ln(1/\delta)$$

$0 \le X \le 1$

BAD

GOOD

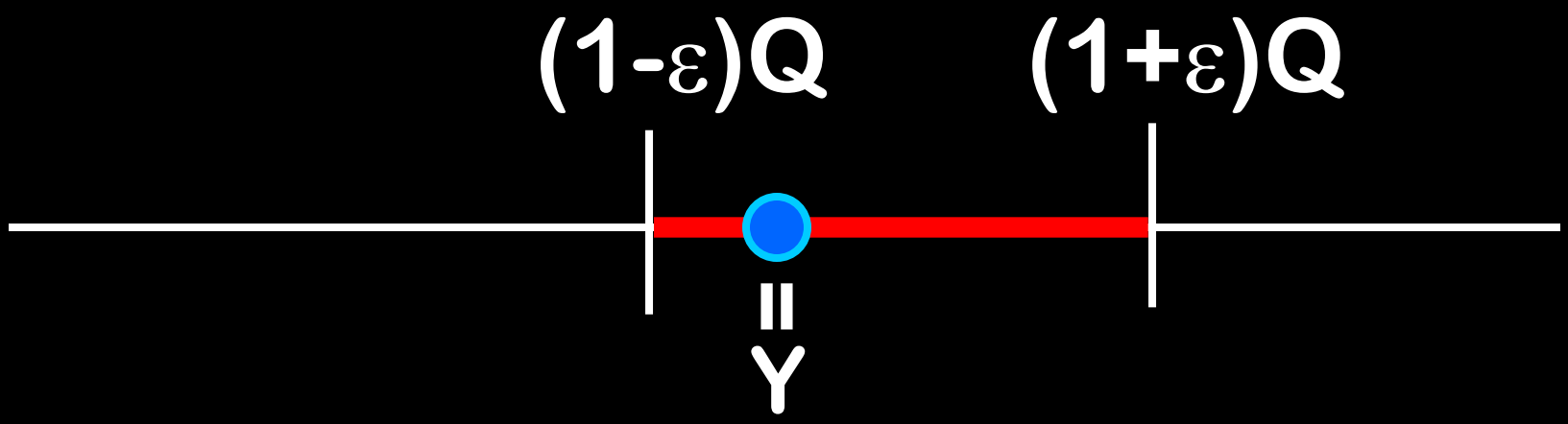
Median “boosting trick”

$$n = \frac{1}{E[X]} \frac{4}{\varepsilon^2}$$

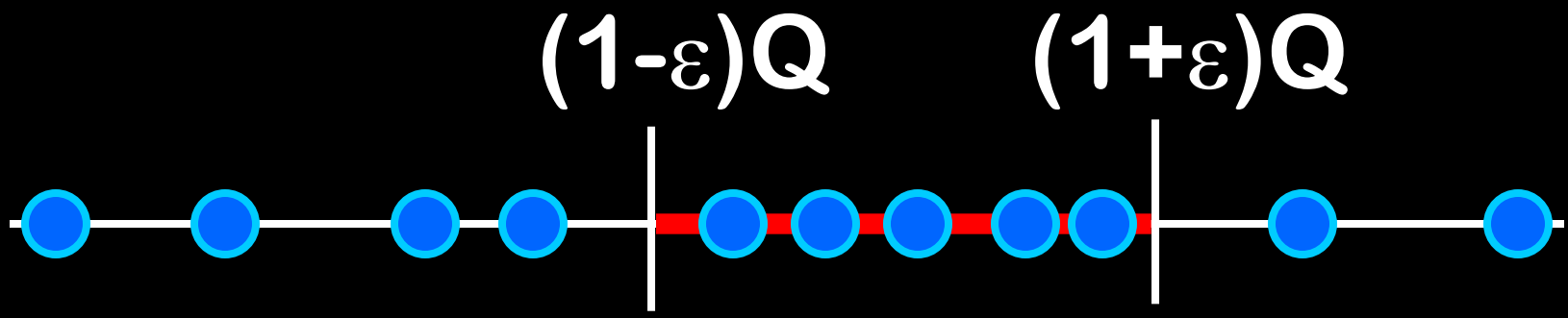
$$Y = \frac{X_1 + X_2 + \ldots + X_n}{n}$$

BY BIENAYME-CHEBYSHEV:

$$P(\bullet \in \text{——}) \geq 3/4$$



Median trick – repeat 2T times



BY BIENAYME-CHEBYSHEV:

$$P(\bullet \in \text{——}) \geq 3/4$$

$\Downarrow$

BY CHERNOFF:

$$P(\text{> T out of 2T}) \geq 1 - e^{-T/4}$$

$\Downarrow$

$$P(\text{median is in}) \geq 1 - e^{-T/4}$$

$$n = \frac{V[X]}{E[X]^2} \frac{32}{\varepsilon^2} \ln(1/\delta)$$

+ median trick

$$n = \frac{1}{E[X]} \frac{3}{\varepsilon^2} \ln(1/\delta)$$

$0 \le X \le 1$

BAD

Creating "approximator" from X

- ε = precision
- δ = confidence

$$n = \Theta\left(\frac{V[X]}{E[X]^2} \frac{1}{\varepsilon^2} \ln(1/\delta) \right)$$

Outline

1. Counting problems

2. Basic tools: Chernoff, Chebyshev

3. Dealing with large quantities
   (the product method)

4. Statistical physics

5. Cooling schedules (our work)

6. More...

(approx) counting $\Leftrightarrow$ sampling

Valleau,Card'72 (physical chemistry), Babai'79 (for matchings and colorings), Jerrum,Valiant,V.Vazirani'86

the outcome of the JVV reduction:

random variables: $X_1 \; X_2 \ldots X_t$
such that

1) $E[X_1 X_2 \ldots X_t]$ = "WANTED"

2) the X_i are easy to estimate

squared coefficient of variation (SCV)

$$\frac{V[X_i]}{E[X_i]^2} = O(1)$$

(approx) counting $\Leftrightarrow$ sampling

1) $E[X_1 X_2 \dots X_t]$ = "WANTED"

2) the X_i are easy to estimate

$$\frac{V[X_i]}{E[X_i]^2} = O(1)$$

Theorem (Dyer-Frieze'91)

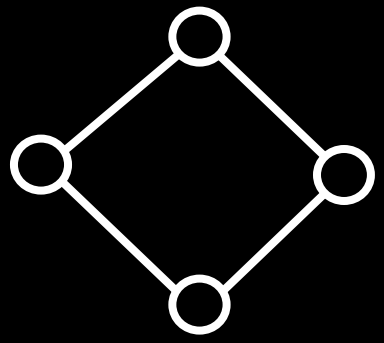
$O(t^2/\varepsilon^2)$ samples ($O(t/\varepsilon^2)$ from each X_i)

give

$1\pm\varepsilon$ estimator of "WANTED" with prob$\geq$3/4

JVV for independent sets

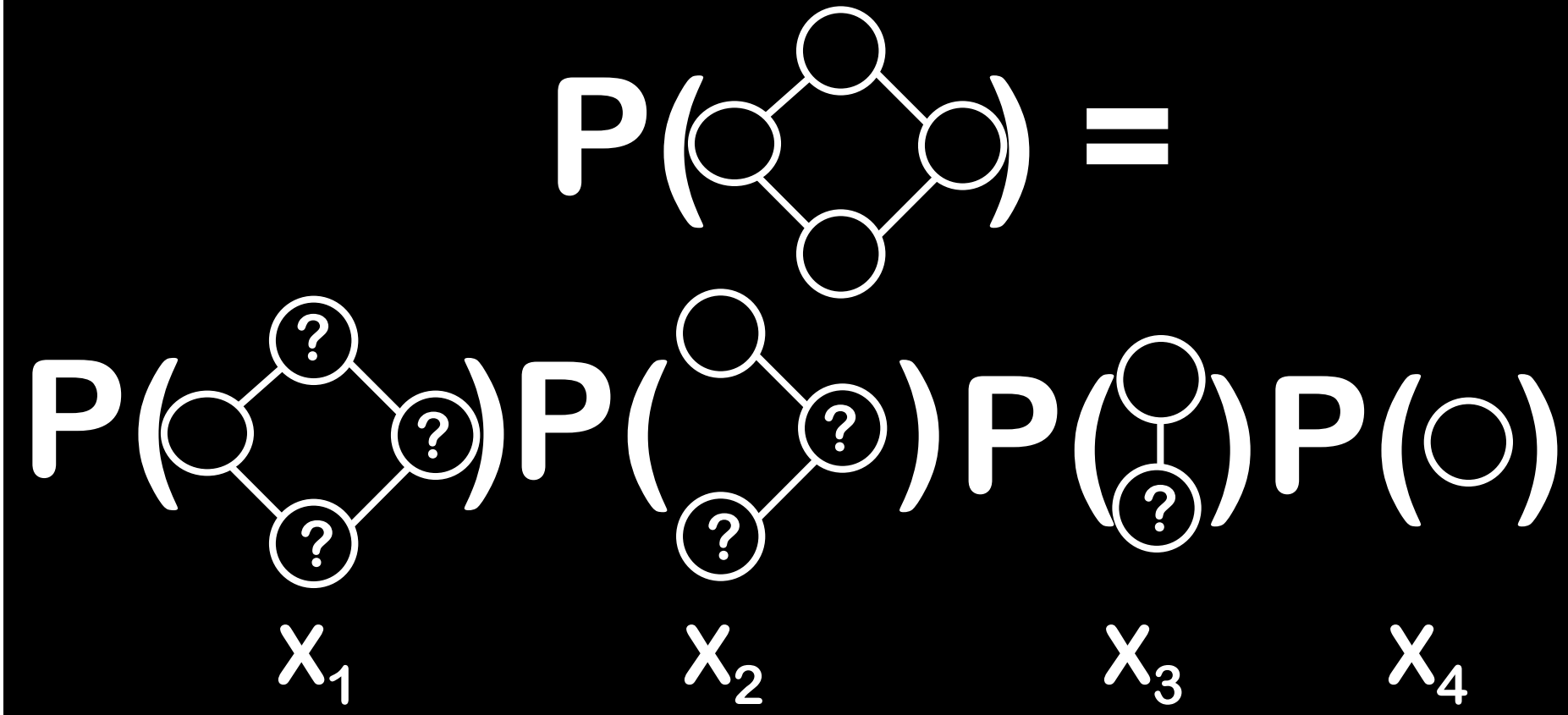
GOAL: given a graph G, estimate the number of independent sets of G



$$\#\text{ independent sets} = \frac{1}{P(\;\;)}$$

JVV for independent sets

$P(A \cap B) = P(A)P(B|A)$



$X_i \in [0,1]$ and $E[X_i] \geq \frac{1}{2} \Rightarrow \frac{V[X_i]}{E[X_i]^2} = O(1)$

JVV for independent sets

$P(A \cap B) = P(A)P(B|A)$

(approx) counting ⇔ sampling

Valleau,Card'72 (physical chemistry), Babai'79 (for matchings and colorings), Jerrum,Valiant,V.Vazirani'86

the outcome of the JVV reduction:

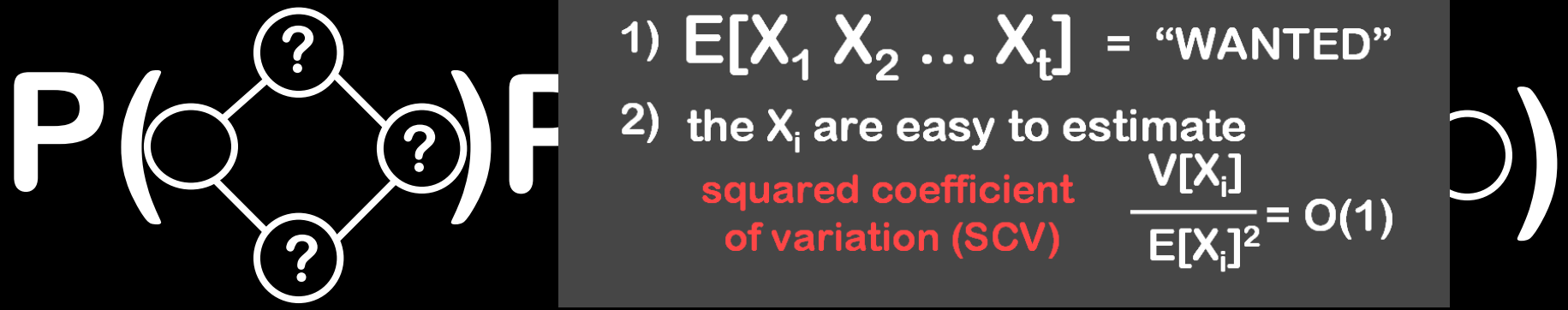
random variables: $X_1 \; X_2 \ldots X_t$

such that

1) $E[X_1 \, X_2 \ldots X_t]$ = "WANTED"

2) the X_i are easy to estimate

squared coefficient of variation (SCV) $\frac{V[X_i]}{E[X_i]^2} = O(1)$

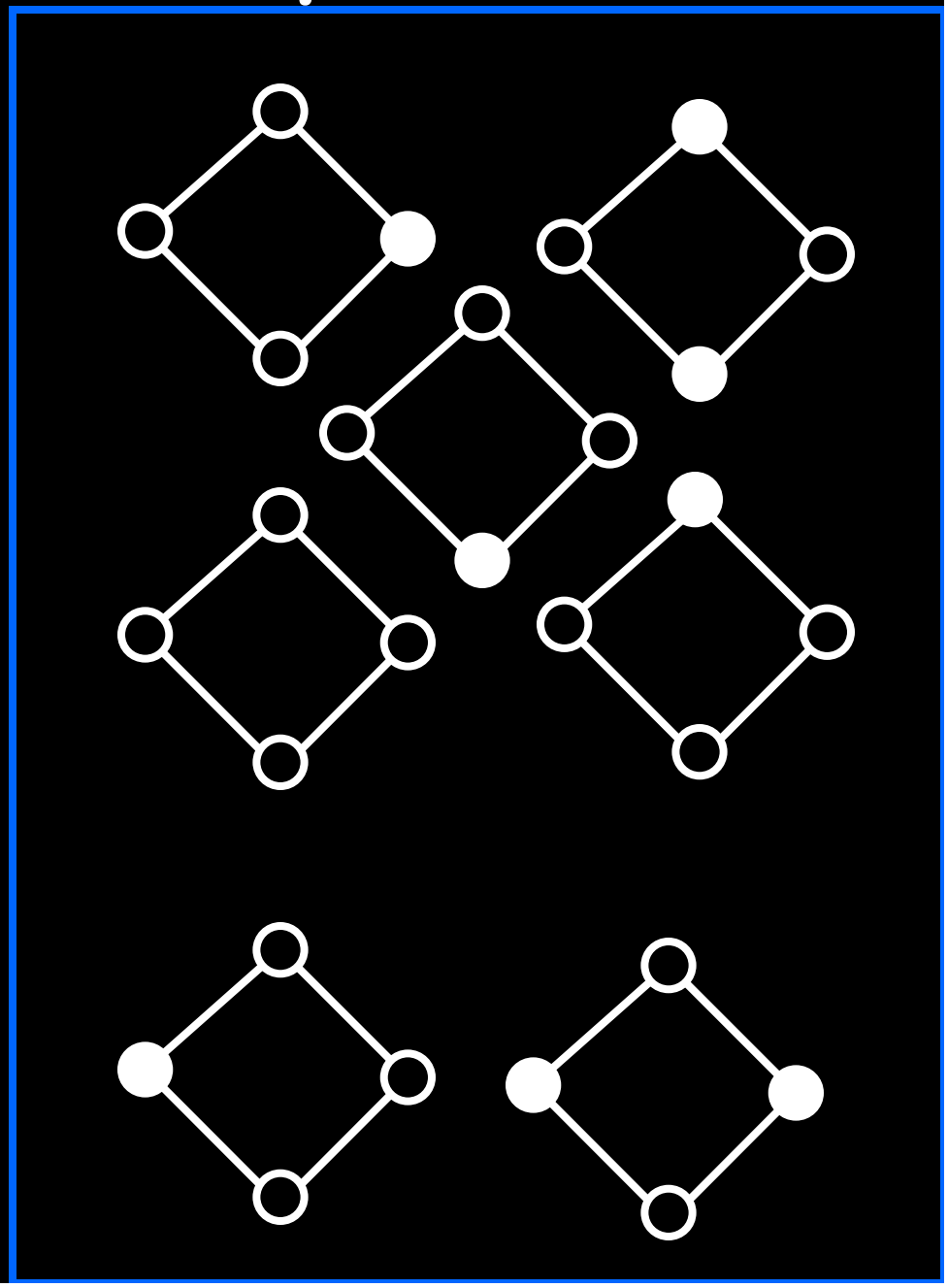


X_1 X_2 X_3 X_4

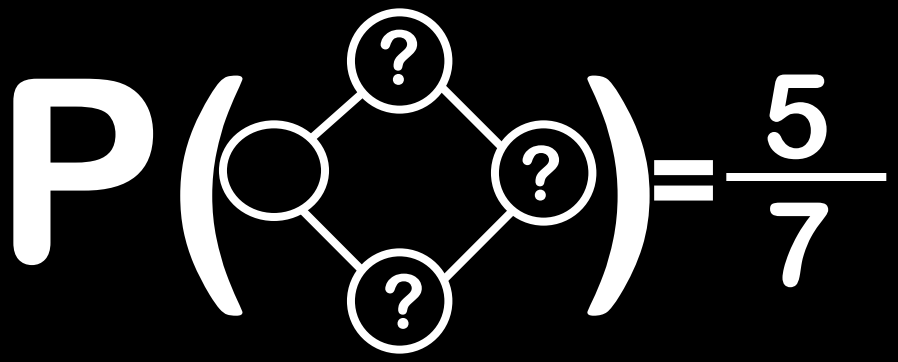
$X_i \in [0,1]$ and $E[X_i] \geq ½ \quad \Rightarrow \quad \frac{V[X_i]}{E[X_i]^2} = O(1)$

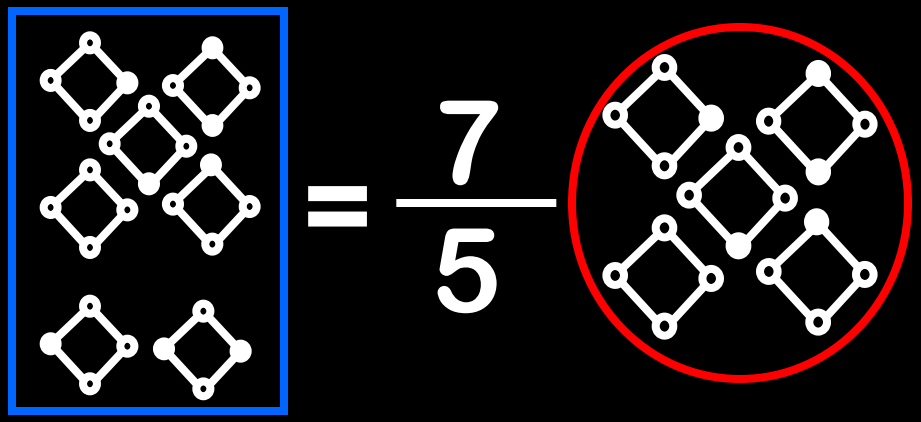
Self-reducibility for independent sets

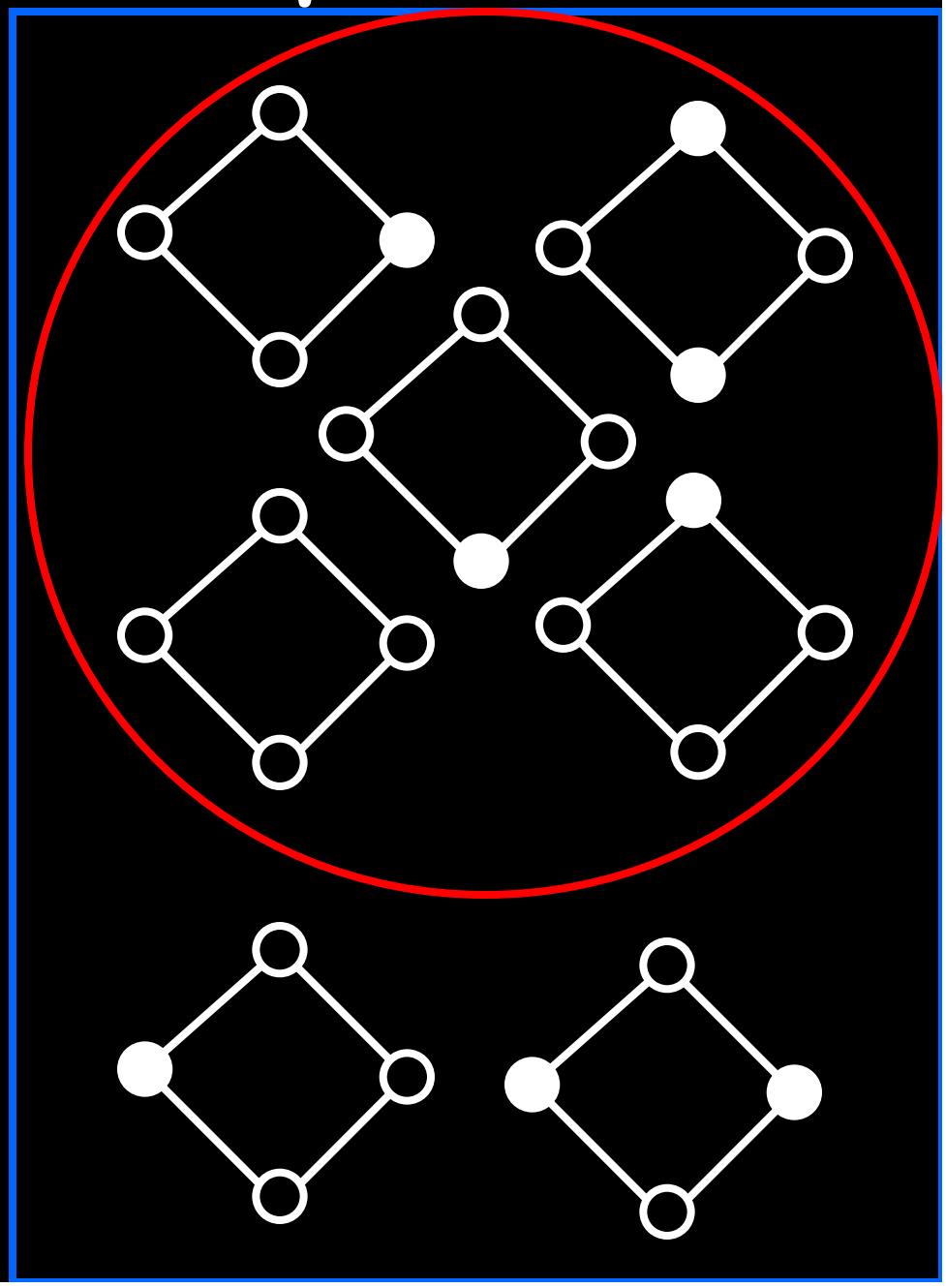
$$P\left(\begin{array}{c}?\\ ?\quad ?\end{array}\right)=\frac{5}{7}$$



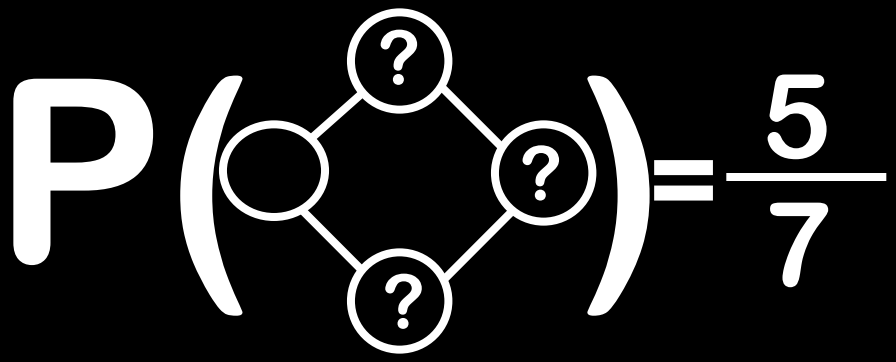
Self-reducibility for independent sets

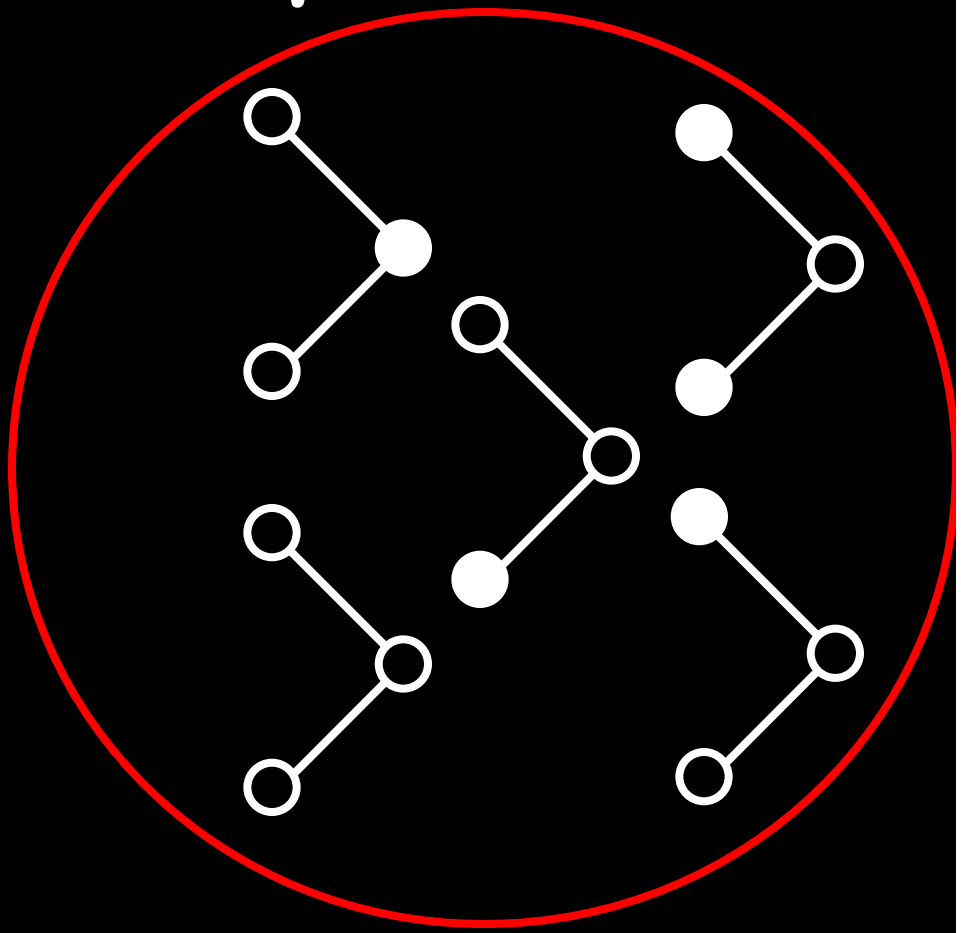


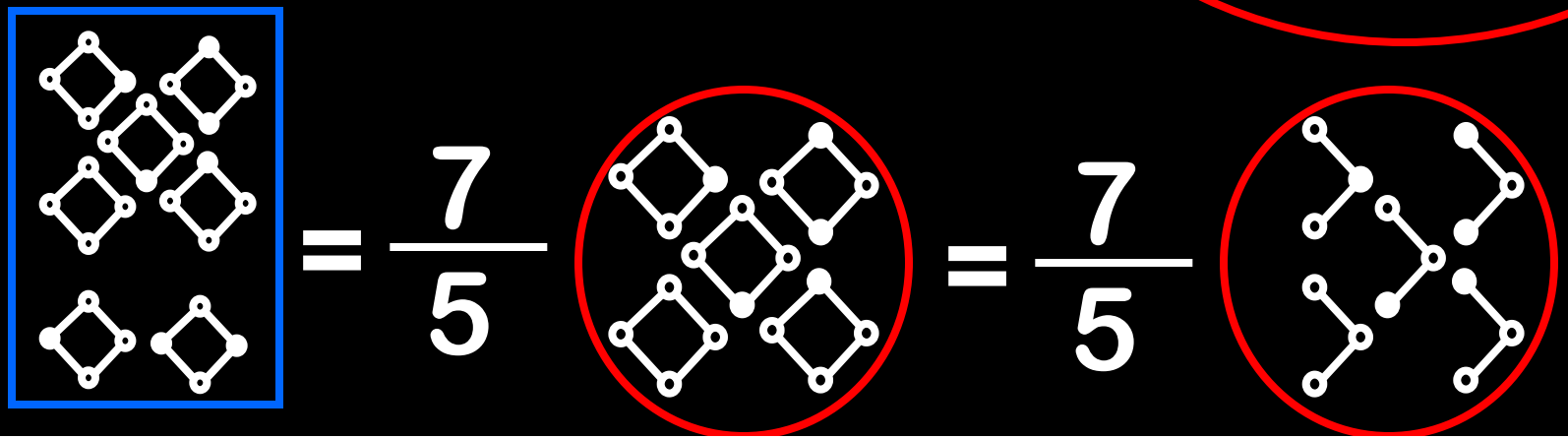




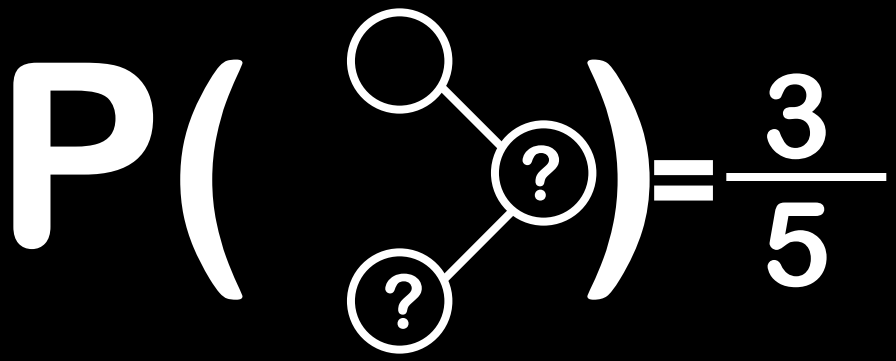
Self-reducibility for independent sets

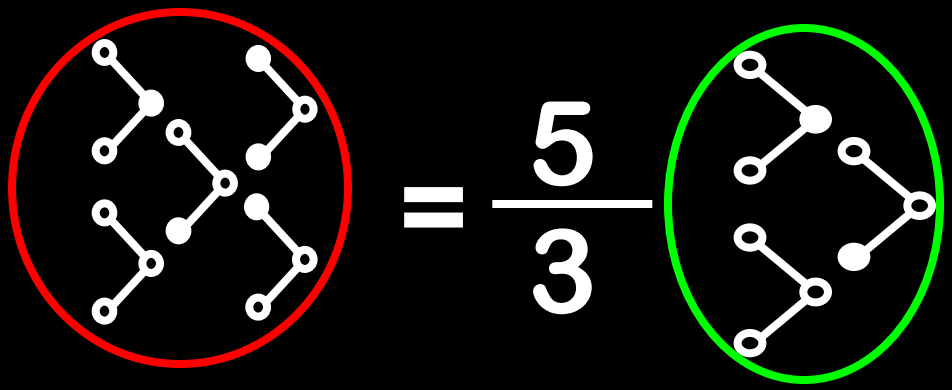


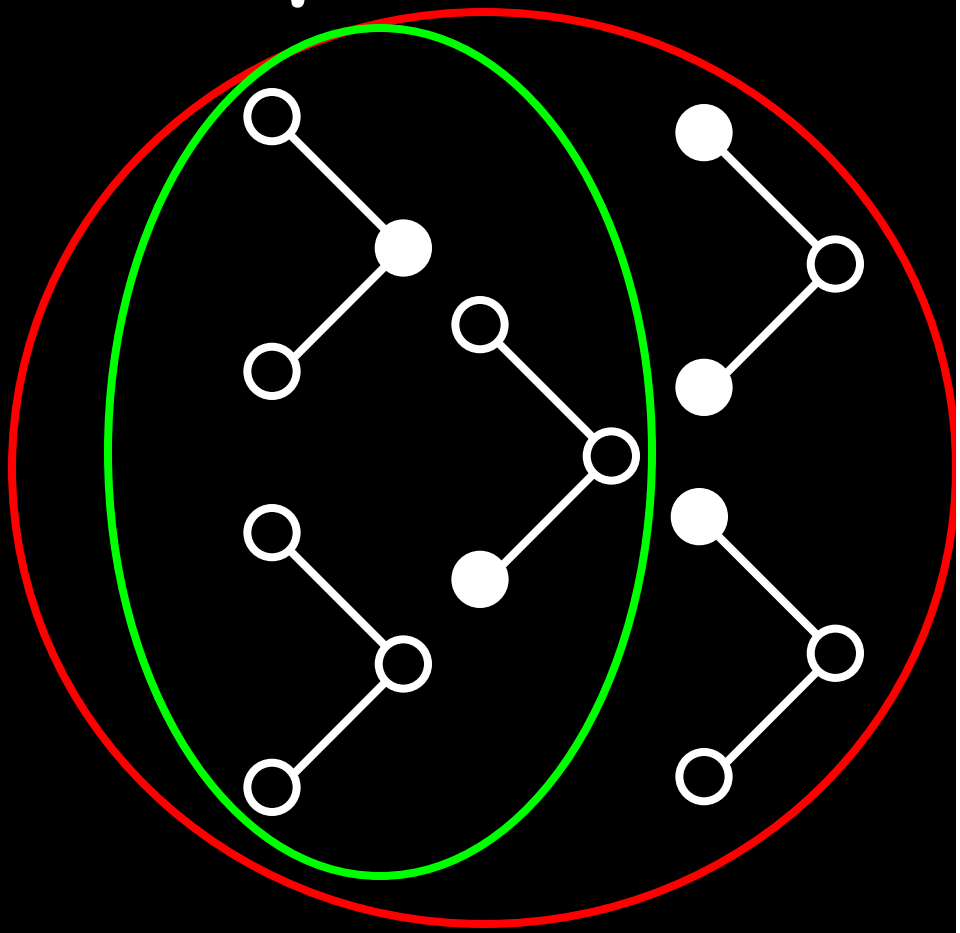




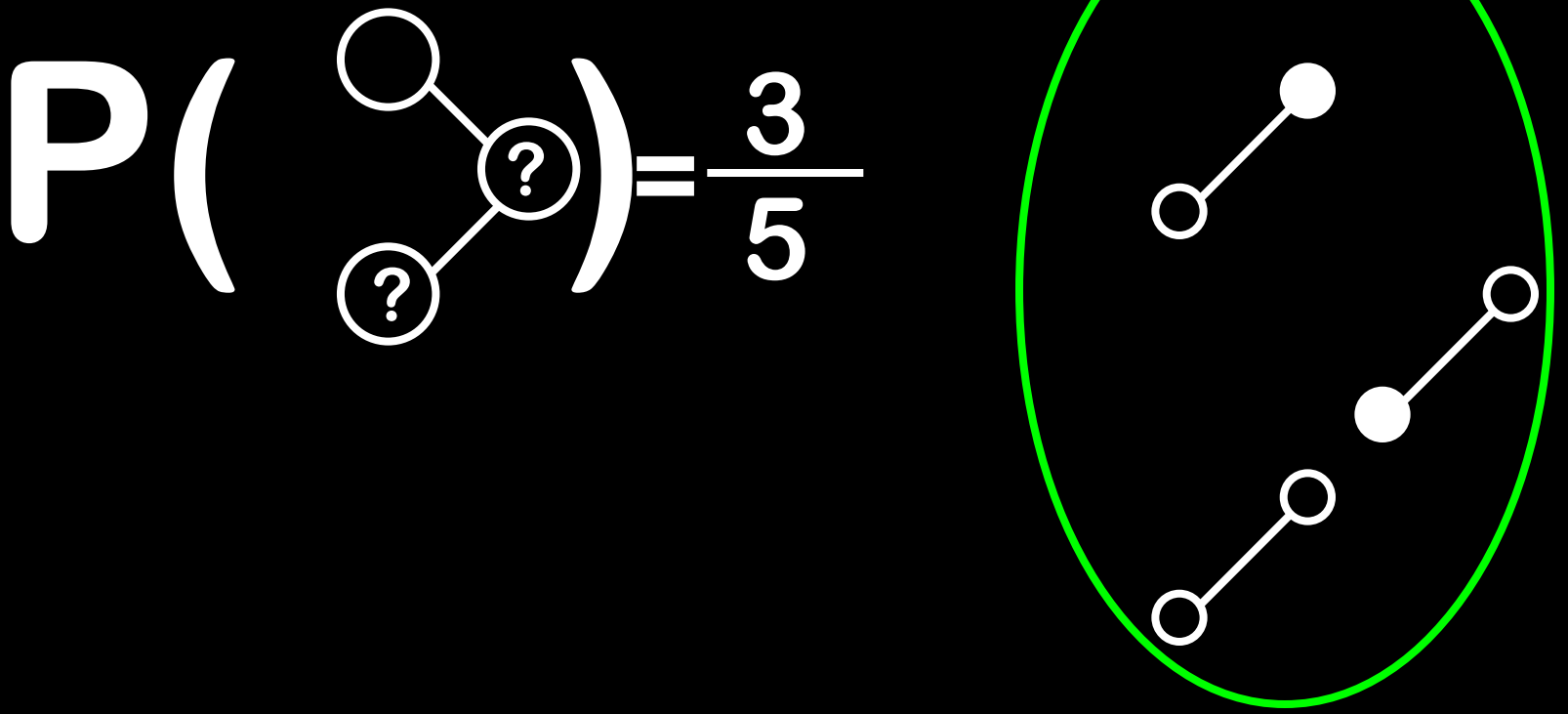
Self-reducibility for independent sets

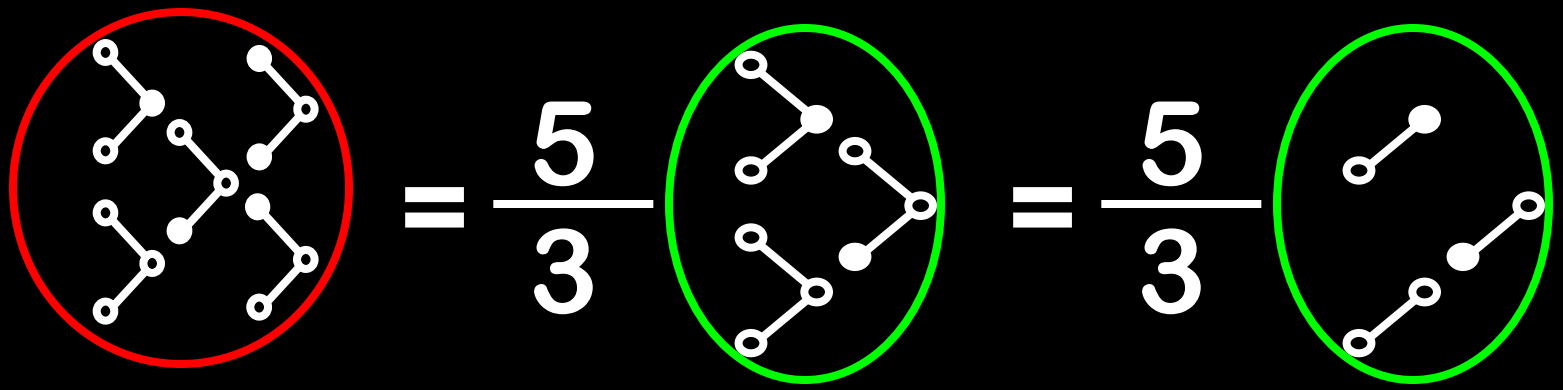




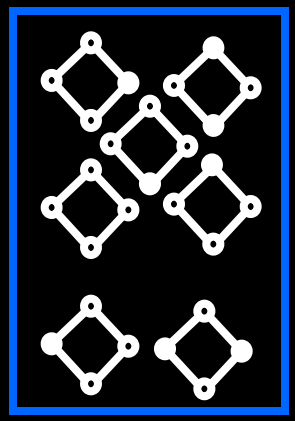


Self-reducibility for independent sets

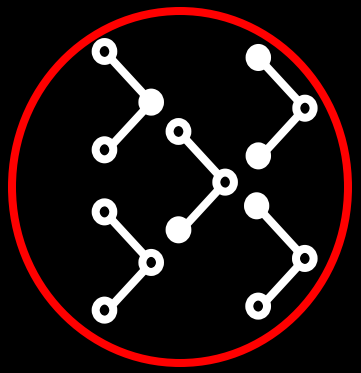




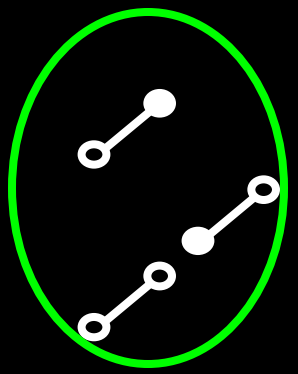
Self-reducibility for independent sets



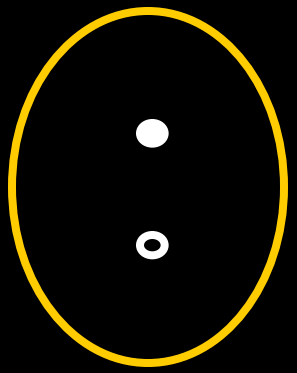
$= \frac{7}{5}$



$= \frac{7}{5}\,\frac{5}{3}$

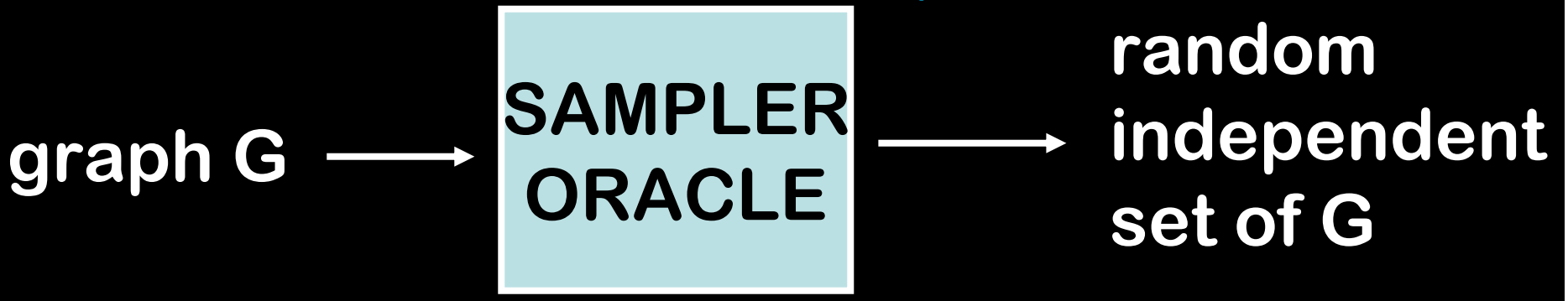


$= \frac{7}{5}\,\frac{5}{3}\,\frac{3}{2}$



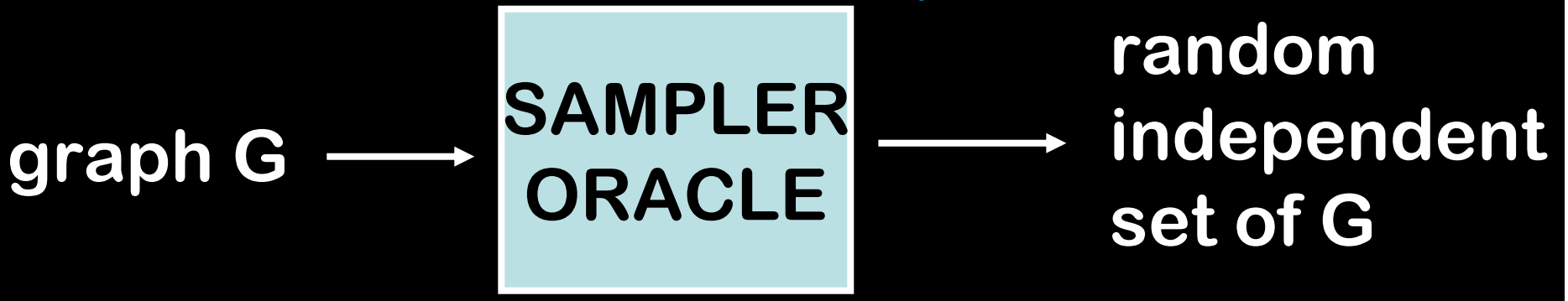
$= 7$

JVV: If we have a sampler oracle:



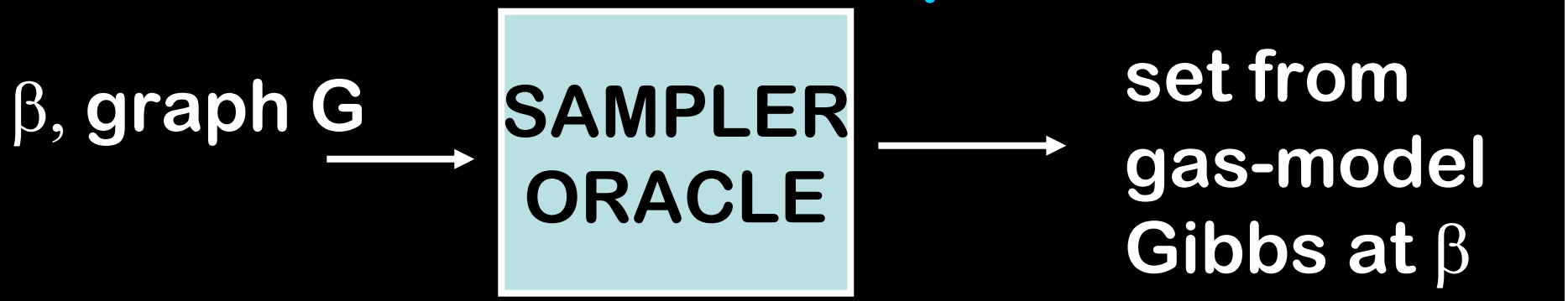
then FPRAS using $O(n^2)$ samples.

JVV: If we have a sampler oracle:



then FPRAS using $O(n^2)$ samples.

ŠVV: If we have a sampler oracle:



then FPRAS using $O^*(n)$ samples.

Application – independent sets

$O^*(|V|)$ samples suffice for counting

Cost per sample (Vigoda'01,Dyer-Greenhill'01)
time = $O^*(|V|)$ for graphs of degree ≤ 4.

Total running time:

$O^* (|V|^2)$.

Other applications

matchings $O^*(n^2m)$

(using Jerrum, Sinclair'89)

spin systems:

Ising model $O^*(n^2)$ for $\beta<\beta_c$

(using Marinelli, Olivieri'95)

k-colorings $O^*(n^2)$ for $k>2\Delta$

(using Jerrum'95)

total running time

Outline

1. Counting problems

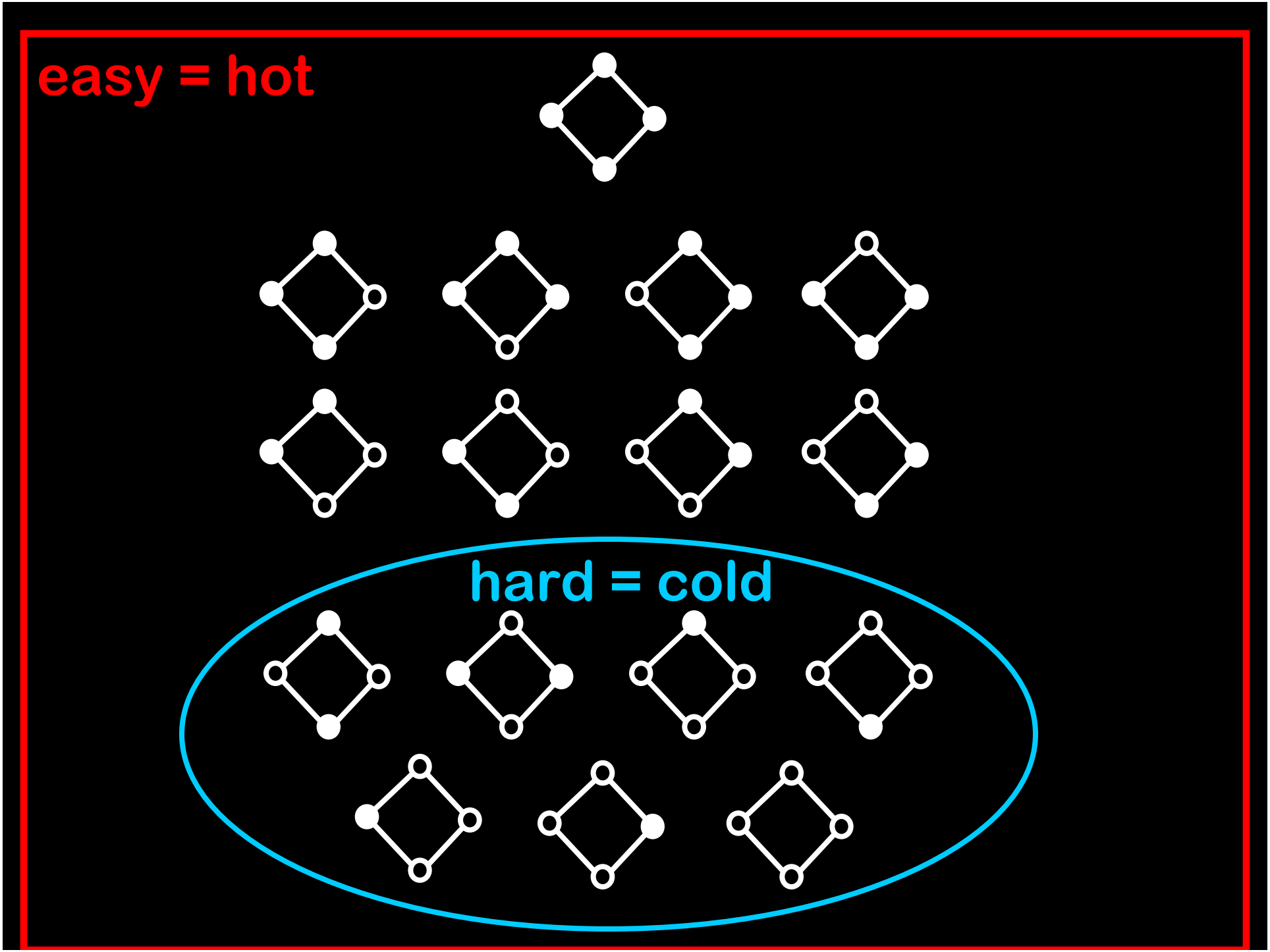
2. Basic tools: Chernoff, Chebyshev

3. Dealing with large quantities
   (the product method)

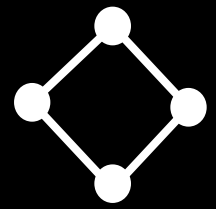
4. Statistical physics

5. Cooling schedules (our work)

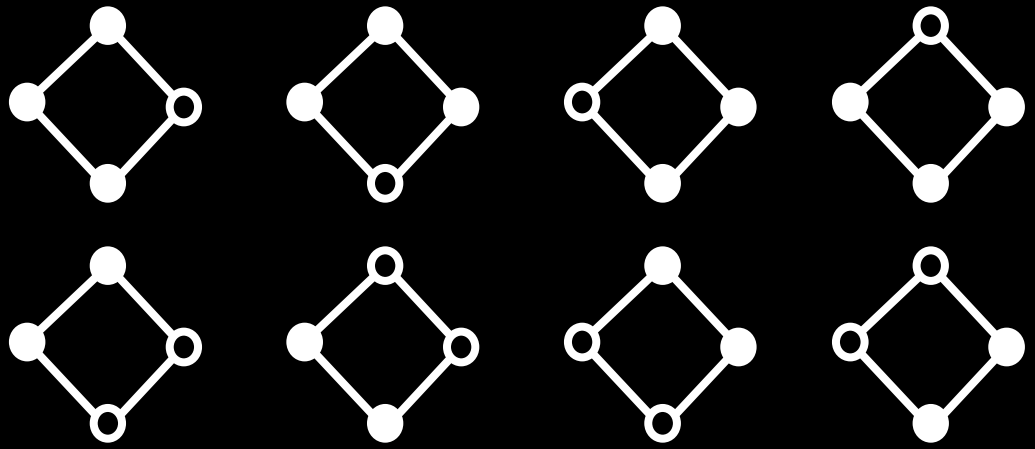
6. More…


easy = hot
hard = cold

Hamiltonian

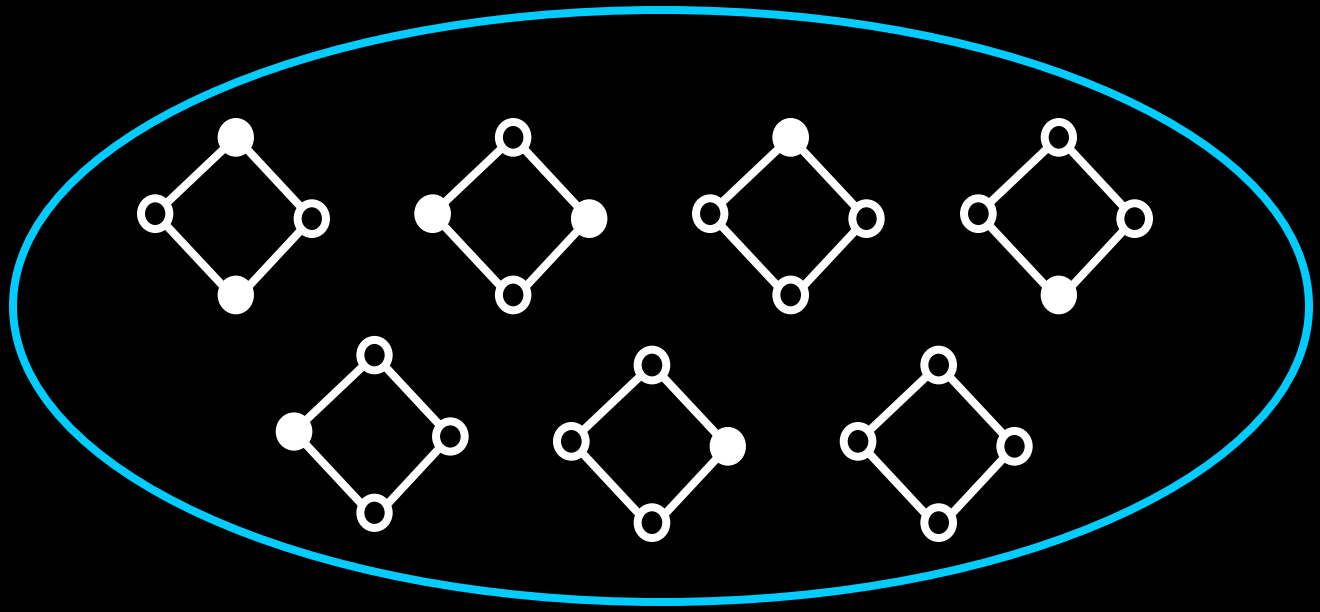


4



2

1



0

Big set = Ω

Hamiltonian

$H : \Omega \to \{0,\ldots,n\}$

Goal: estimate $|H^{-1}(0)|$

$|H^{-1}(0)| = E[X_1] \ldots E[X_t]$

Distributions between hot and cold

β = inverse temperature

$\beta = 0 \Rightarrow$ hot $\Rightarrow$ uniform on Ω

$\beta = \infty \Rightarrow$ cold $\Rightarrow$ uniform on $H^{-1}(0)$

$$\mu_\beta(x) \propto \exp(-H(x)\beta)$$

(Gibbs distributions)

Distributions between hot and cold

$$\mu_\beta(x) \propto \exp(-H(x)\beta)$$

$$\mu_\beta(x) = \frac{\exp(-H(x)\beta)}{Z(\beta)}$$

Normalizing factor = partition function

$$Z(\beta) = \sum_{x \in \Omega} \exp(-H(x)\beta)$$

Partition function

$$Z(\beta) = \sum_{x \in \Omega} \exp(-H(x)\beta)$$

have: $Z(0) = |\Omega|$

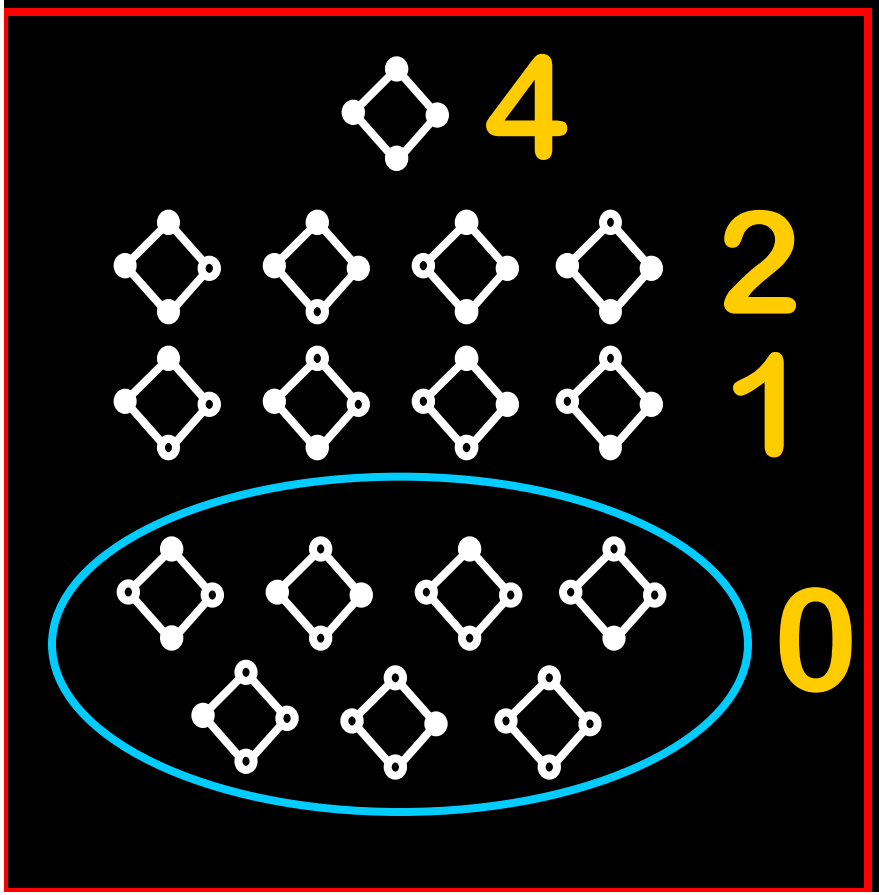
want: $Z(\infty) = |H^{-1}(0)|$

Partition function - example

$$Z(\beta) = \sum_{x \in \Omega} \exp(-H(x)\beta)$$

have: $Z(0) = |\Omega|$

want: $Z(\infty) = |H^{-1}(0)|$



$$Z(\beta) = 1\, e^{-4.\beta} + 4\, e^{-2.\beta} + 4\, e^{-1.\beta} + 7\, e^{-0.\beta}$$

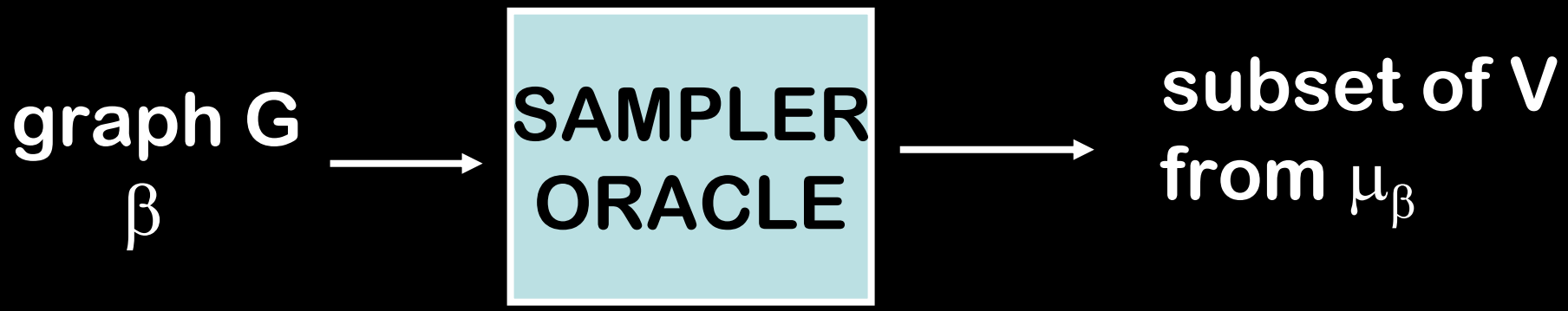
$Z(0) = 16$

$Z(\infty) = 7$

Assumption:

we have a sampler oracle for μ_β

$$\mu_\beta(x) = \frac{\exp(-H(x)\beta)}{Z(\beta)}$$



Assumption:

we have a sampler oracle for μ_β

$$\mu_\beta(x) = \frac{\exp(-H(x)\beta)}{Z(\beta)}$$

$$W \sim \mu_\beta$$

Assumption:

we have a sampler oracle for μ_β

$$\mu_\beta(x) = \frac{\exp(-H(x)\beta)}{Z(\beta)}$$

$$W \sim \mu_\beta \rightarrow X = \exp(H(W)(\beta - \alpha))$$

Assumption:

we have a sampler oracle for μ_β

$$\mu_\beta(x) = \frac{\exp(-H(x)\beta)}{Z(\beta)}$$

$$W \sim \mu_\beta \rightarrow X = \exp(H(W)(\beta - \alpha))$$

can obtain the following ratio:

$$E[X] = \sum_{s \in \Omega} \mu_\beta(s)\, X(s) = \frac{Z(\alpha)}{Z(\beta)}$$

Our goal restated

Partition function

$$Z(\beta) = \sum_{x \in \Omega} \exp(-H(x)\beta)$$

Goal: estimate $Z(\infty) = |H^{-1}(0)|$

$$Z(\infty) = \frac{Z(\beta_1)}{Z(\beta_0)} \frac{Z(\beta_2)}{Z(\beta_1)} \cdots \frac{Z(\beta_t)}{Z(\beta_{t-1})} Z(0)$$

$$\beta_0 = 0 < \beta_1 < \beta_2 < \ldots < \beta_t = \infty$$

Our goal restated

$$Z(\infty) = \frac{Z(\beta_1)}{Z(\beta_0)} \frac{Z(\beta_2)}{Z(\beta_1)} \cdots \frac{Z(\beta_t)}{Z(\beta_{t-1})} Z(0)$$

Cooling schedule:

$$\beta_0 = 0 < \beta_1 < \beta_2 < \ldots < \beta_t = \infty$$

How to choose the cooling schedule?

minimize length, while satisfying

$$\frac{V[X_i]}{E[X_i]^2} = O(1) \qquad E[X_i] = \frac{Z(\beta_i)}{Z(\beta_{i-1})}$$

Our goal restated

$$Z(\infty) = \frac{Z(\beta_1)}{Z(\beta_0)} \cdot \frac{Z(\beta_2)}{Z(\beta_1)} \cdots \frac{Z(\beta_n)}{Z(\beta_{n-1})}$$

Cooling schedule:

$$\beta_0 = 0 < \beta_1 < \cdots$$

How to choose the

minimize leng

$$\frac{V[X_i]}{E[X_i]^2} = O(1) \qquad E[X_i] = \frac{Z(\beta_i)}{Z(\beta_{i-1})}$$

Assumption:

we have a sampler oracle for μ_β

$$\mu_\beta(x) = \frac{\exp(-H(x)\beta)}{Z(\beta)}$$

$$W \sim \mu_\beta \rightarrow X = \exp(H(W)(\beta - \alpha))$$

can obtain the following ratio:

$$E[X] = \sum_{s \in \Omega} \mu_\beta(s)\, X(s) = \frac{Z(\alpha)}{Z(\beta)}$$

Outline

1. Counting problems

2. Basic tools: Chernoff, Chebyshev

3. Dealing with large quantities
   (the product method)

4. Statistical physics

5. Cooling schedules (our work)

6. More...

Parameters: A and n

$$Z(\beta) = \sum_{x \in \Omega} \exp(-H(x)\beta)$$

$$Z(0) = A$$

$$H:\Omega \rightarrow \{0,\ldots,n\}$$

$$Z(\beta) = \sum_{k=0}^{n} a_k \, e^{-\beta k}$$

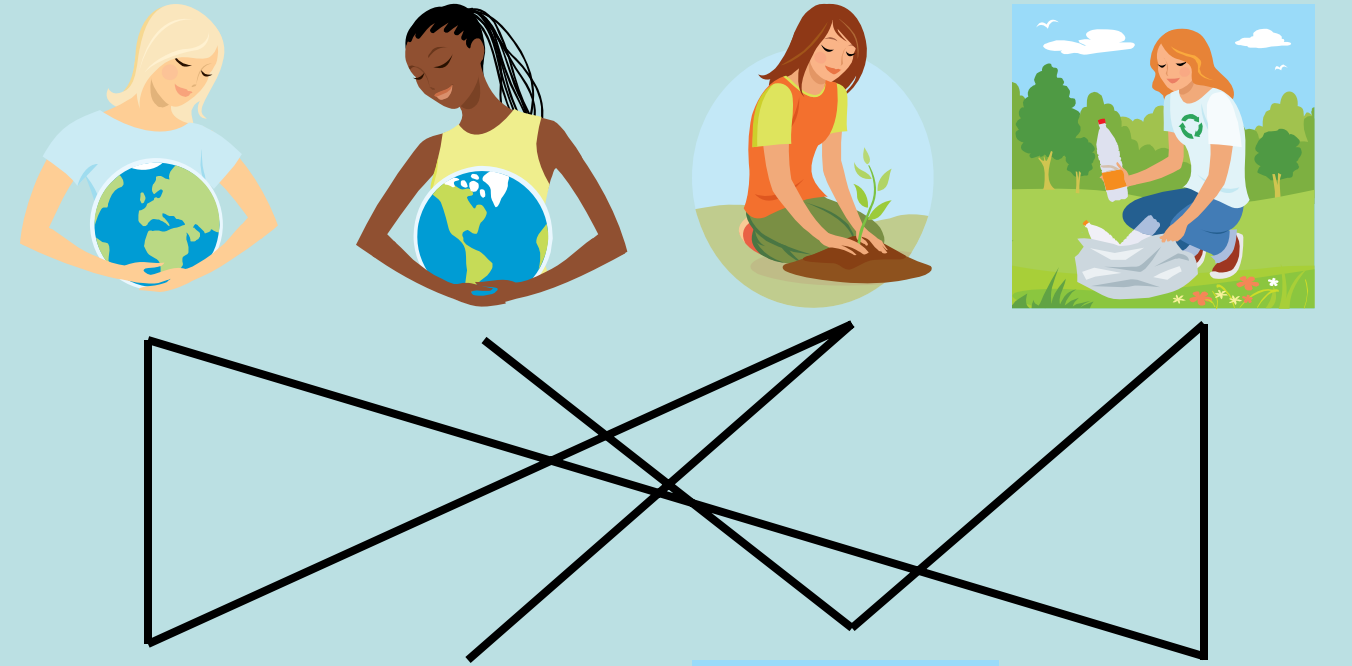
$$a_k = |H^{-1}(k)|$$

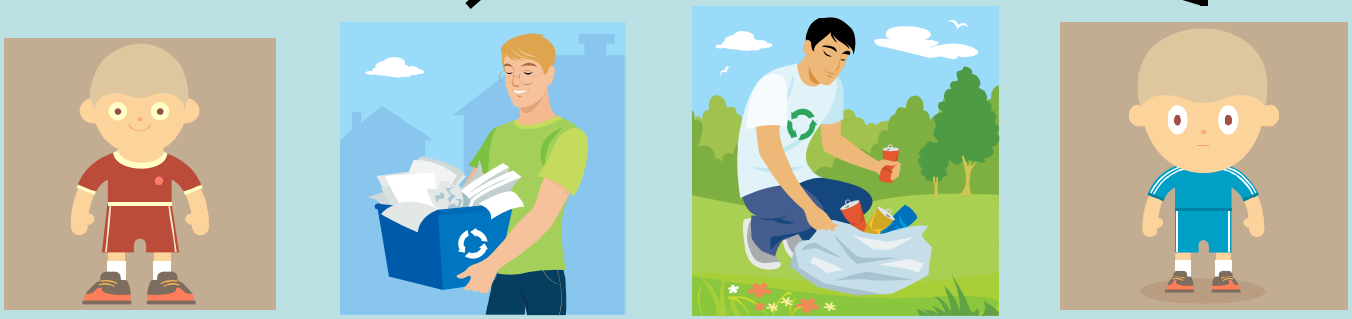
Parameters

$Z(0) = A$ $\quad H:\Omega \to \{0,\ldots,n\}$

	A	n
independent sets	2^V	E
matchings	$\approx V!$	V
perfect matchings	V!	V
k-colorings	k^V	E

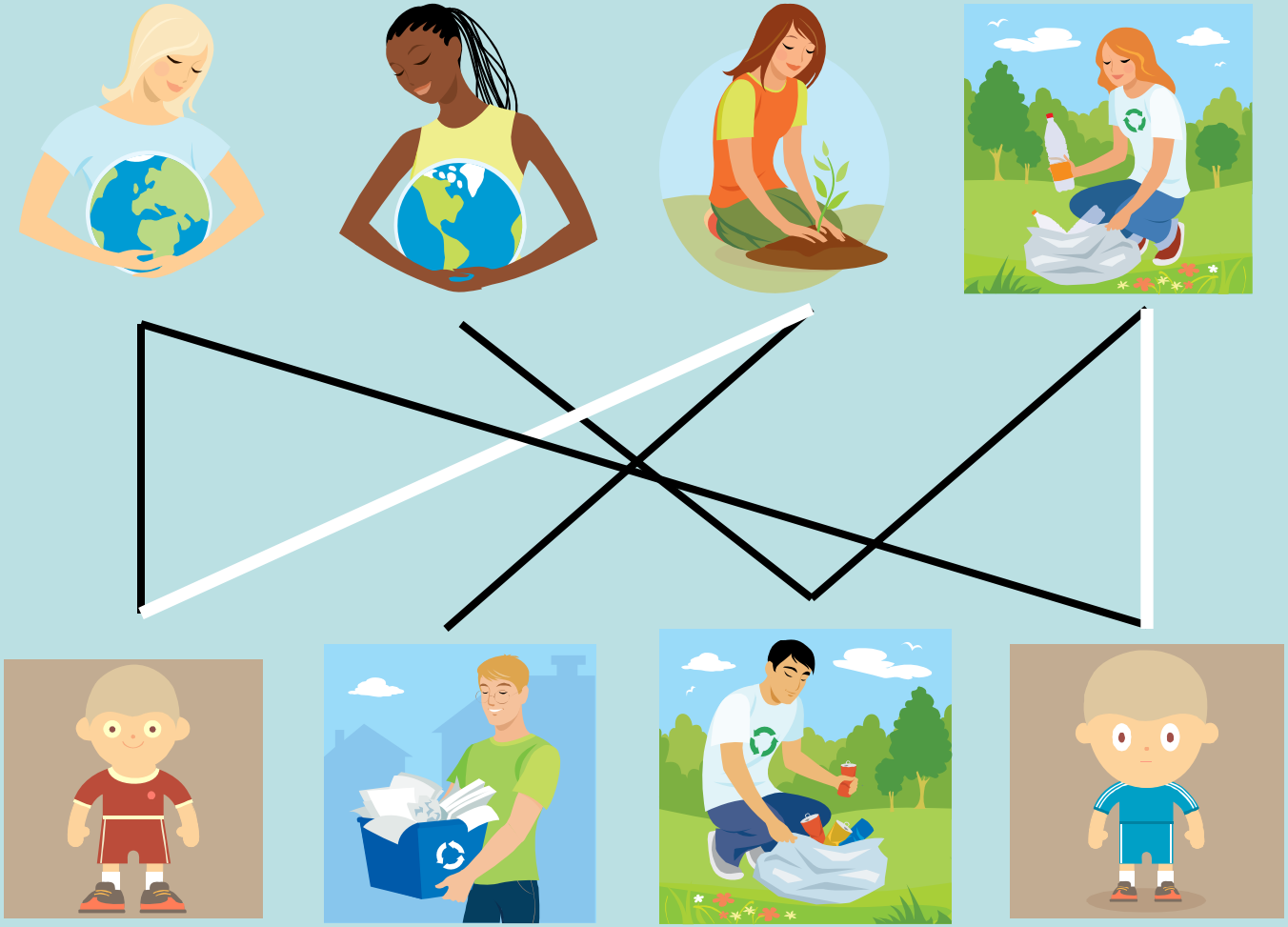
Parameters





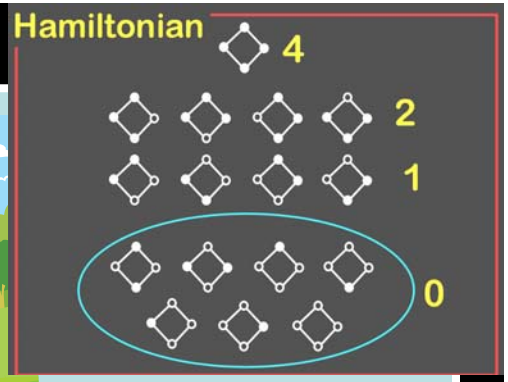
matchings = # ways of marrying them so that no unhappy couple

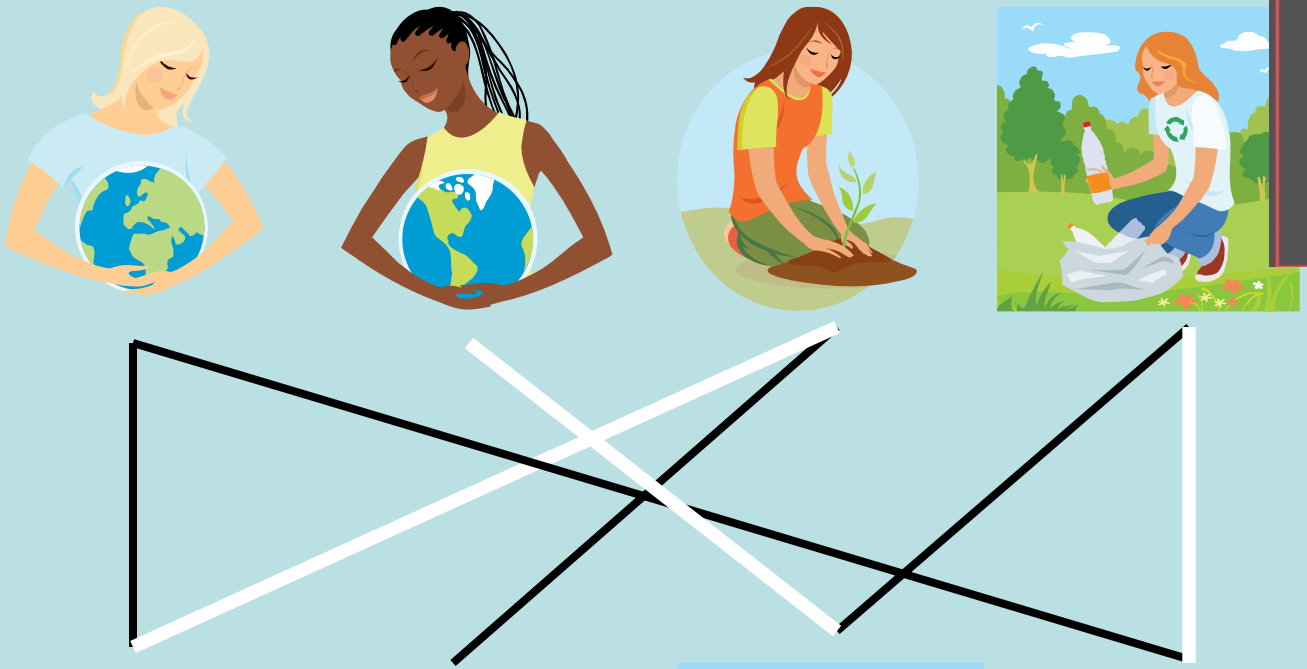
Parameters

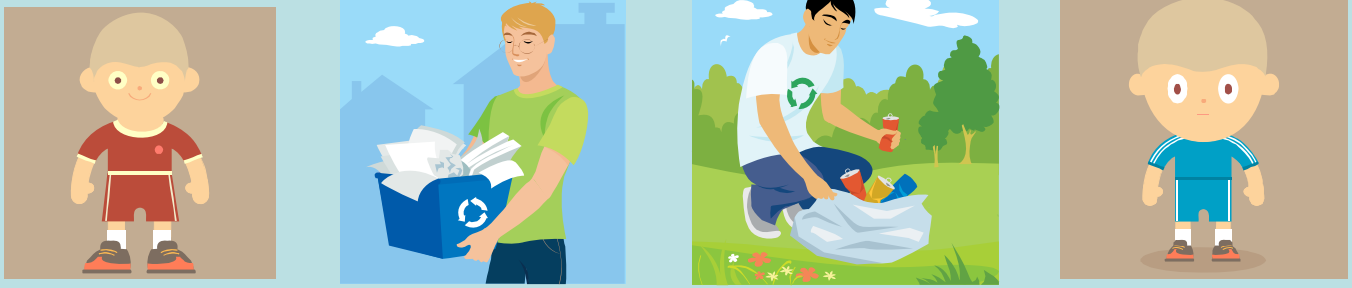


matchings = # ways of marrying them so that no unhappy couple

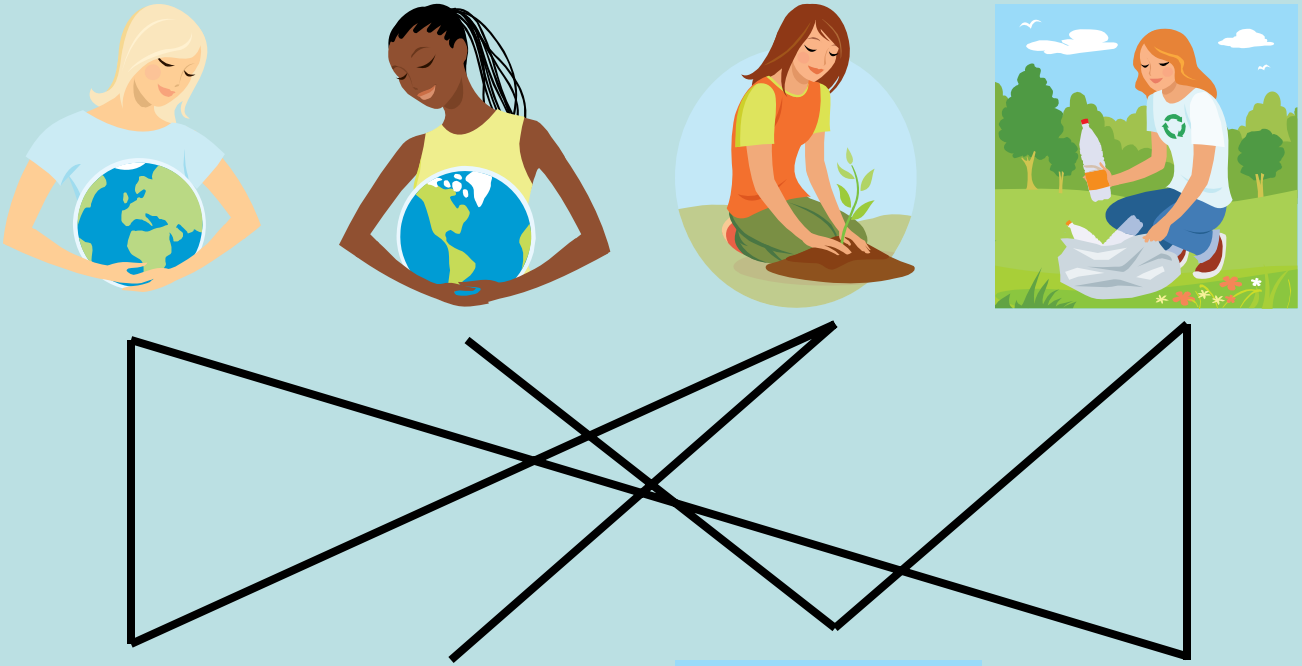
Parameters

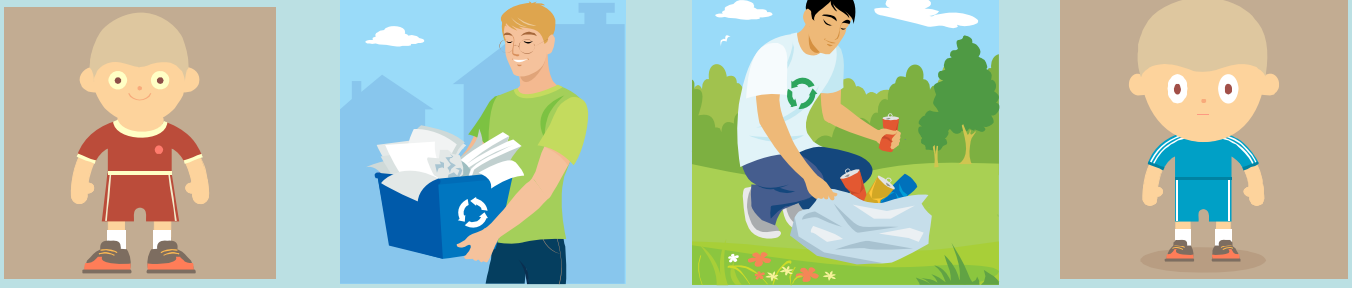






matchings = # ways of marrying them so that no unhappy couple





marry ignoring “compatibility”

hamiltonian = number of unhappy couples

Parameters

$Z(0) = A$ $\quad H:\Omega \to \{0,\ldots,n\}$

	A	n
independent sets	2^V	E
matchings	$\approx V!$	V
perfect matchings	V!	V
k-colorings	k^V	E

Previous cooling schedules

$Z(0) = A$ $\quad H:\Omega \to \{0,\ldots,n\}$

$\beta_0 = 0 < \beta_1 < \beta_2 < \ldots < \beta_t = \infty$

"Safe steps"

$\beta \to \beta + 1/n$

$\beta \to \beta\,(1 + 1/\ln A)$

$\ln A \to \infty$

(Bezáková,Štefankovič, Vigoda,V.Vazirani'06)

Cooling schedules of length

$O(n \ln A)$

$O((\ln n)(\ln A))$

(Bezáková,Štefankovič, Vigoda,V.Vazirani'06)

Previous cooling s

$Z(0) = A$ H

$\beta_0 = 0 < \beta_1 < \beta$

"Safe steps"

$\beta \rightarrow \beta + 1/n$

$\beta \rightarrow \beta\,(1 + 1/\ln A)$

$\ln A \rightarrow \infty$

(Bezáková,Štefankovič, Vigoda,V.Vazirani'06)

Cooling schedules of length

$O(n \ln A)$

$O((\ln n)(\ln A))$

(Bezáková,Štefankovič, Vigoda,V.Vazirani'06)

Our goal restated

$$Z(\infty) = \frac{Z(\beta_1)}{Z(\beta_0)} \frac{Z(\beta_2)}{Z(\beta_1)} \cdots \frac{Z(\beta_t)}{Z(\beta_{t-1})} Z(0)$$

Cooling schedule:

$\beta_0 = 0 < \beta_1 < \beta_2 < \ldots < \beta_t = \infty$

How to choose the cooling schedule?

minimize length, while satisfying

$$\frac{V[X_i]}{E[X_i]^2} = O(1) \qquad E[X_i] = \frac{Z(\beta_i)}{Z(\beta_{i-1})}$$

"Safe steps"

$\beta \to \beta + 1/n$

$\beta \to \beta\,(1 + 1/\ln A)$

$\ln A \to \infty$

(Bezáková, Štefankovič, Vigoda, V. Vazirani'06)

$$W \sim \mu_\beta \to X = \exp(H(W)(\beta - \alpha))$$

$$Z(\beta) = \sum_{k=0}^{n} a_k\, e^{-\beta k}$$

$$1/e \le X \le 1$$

$$\frac{V[X]}{E[X]^2} \le \frac{1}{E[X]} \le e$$

"Safe steps"

$\beta \to \beta + 1/n$

$\beta \to \beta\,(1 + 1/\ln A)$

$\ln A \to \infty$

(Bezáková,Štefankovič, Vigoda,V.Vazirani'06)

$$W \sim \mu_\beta \rightarrow X \equiv \exp(H(W)(\beta - \alpha))$$

$$Z(\beta) = \sum_{k=0}^{n} a_k\, e^{-\beta k}$$

$Z(\infty) = a_0 \geq 1$

$Z(\ln A) \leq a_0 + 1$

$E[X] \geq 1/2$

“Safe steps”

$\beta \to \beta + 1/n$

$\beta \to \beta\ (1 + 1/\ln A)$

$\ln A \to \infty$

(Bezáková,Štefankovič, Vigoda,V.Vazirani’06)

$$W \sim \mu_\beta \rightarrow X \equiv \exp(H(W)(\beta - \alpha))$$

$$Z(\beta) = \sum_{k=0}^{n} a_k e^{-\beta k}$$

$$E[X] \geq 1/2e$$

Previous cooling schedules

1/n, 2/n, 3/n, …. , (ln A)/n, …. , ln A

Our goal restated

$$Z(\infty) = \frac{Z(\beta_1)}{Z(\beta_0)} \frac{Z(\beta_2)}{Z(\beta_1)} \cdots \frac{Z(\beta_t)}{Z(\beta_{t-1})} Z(0)$$

Cooling schedule:

$$\beta_0 = 0 < \beta_1 < \beta_2 < \ldots < \beta_t = \infty$$

How to choose the cooling schedule?

minimize length, while satisfying

$$\frac{V[X_i]}{E[X_i]^2} = O(1) \qquad E[X_i] = \frac{Z(\beta_i)}{Z(\beta_{i-1})}$$

"Safe steps"

$\beta \rightarrow \beta + 1/n$

$\beta \rightarrow \beta\,(1 + 1/\ln A)$

$\ln A \rightarrow \infty$

(Bezáková,Štefankovič, Vigoda,V.Vazirani'06)

Cooling schedules of length

O(n ln A)

O((ln n) (ln A))

(Bezáková,Štefankovič, Vigoda,V.Vazirani'06)

No better fixed schedule possible

$Z(0) = A$ $\quad H:\Omega \rightarrow \{0,\ldots,n\}$

THEOREM:

A schedule that works for all

$$Z_a(\beta) = \frac{A}{1+a}\left(1 + a\, e^{-\beta n}\right)$$

(with $a \in [0, A-1]$)

has LENGTH $\geq \Omega((\ln n)(\ln A))$

Parameters

$Z(0) = A$ $\quad\quad$ $H:\Omega \rightarrow \{0,\ldots,n\}$

Our main result:

can get <u>adaptive</u> schedule of length $O^*((\ln A)^{1/2})$

Previously:

<u>non-adaptive</u> schedules of length $\Omega^*(\ln A)$

Related work

can get adaptive schedule of length $O^*((\ln A)^{1/2})$

Lovász-Vempala
Volume of convex bodies in $O^*(n^4)$
schedule of length $O(n^{1/2})$

(non-adaptive cooling schedule, using specific properties of the "volume" partition functions)

Existential part

Lemma:

for every partition function there exists a cooling schedule of length $O^*((\ln A)^{1/2})$

there exists

can get adaptive schedule of length $O^*((\ln A)^{1/2})$

Cooling schedule (definition refresh)

$$Z(\infty) = \frac{Z(\beta_1)}{Z(\beta_0)} \frac{Z(\beta_2)}{Z(\beta_1)} \cdots \frac{Z(\beta_t)}{Z(\beta_{t-1})} Z(0)$$

Cooling schedule:

$$\beta_0 = 0 < \beta_1 < \beta_2 < \ldots < \beta_t = \infty$$

How to choose the cooling schedule?

minimize length, while satisfying

$$\frac{V[X_i]}{E[X_i]^2} = O(1)$$

$$E[X_i] = \frac{Z(\beta_i)}{Z(\beta_{i-1})}$$

Express SCV using partition function

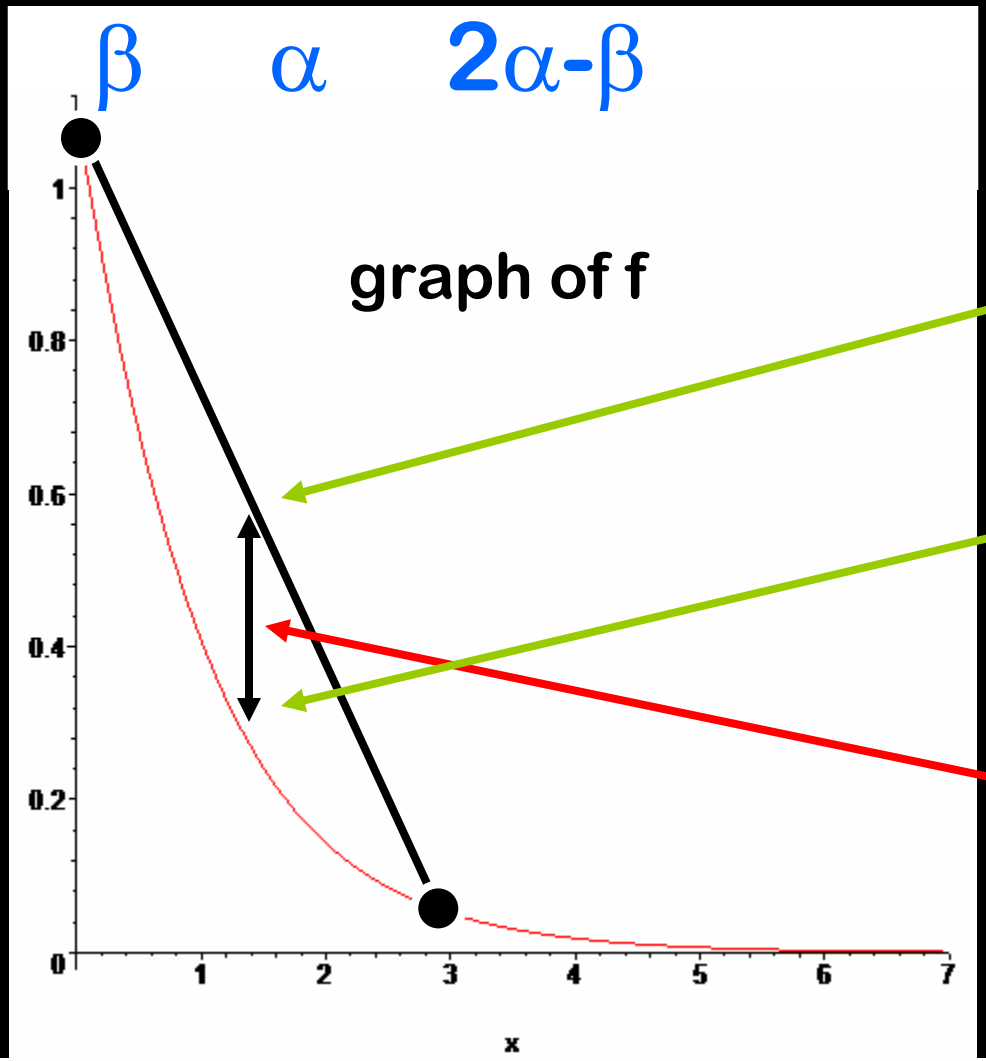
(going from β to α) $E[X] = \frac{Z(\alpha)}{Z(\beta)}$

$W \sim \mu_\beta \rightarrow X = \exp(H(W)(\beta - \alpha))$

$$\frac{E[X^2]}{E[X]^2} = \frac{Z(2\alpha-\beta)\, Z(\beta)}{Z(\alpha)^2} \leq C$$

$$\frac{V[X]}{E[X]^2} + 1$$

$$\frac{E[X^2]}{E[X]^2} = \frac{Z(2\alpha-\beta)\,Z(\beta)}{Z(\alpha)^2} \leq C$$



$$f(\gamma)=\ln Z(\gamma)$$

$$(f(2\alpha-\beta) + f(\beta))/2 \leq (\ln C)/2 + f(\alpha)$$

$$\leq C'=(\ln C)/2$$

Properties of partition functions

$f(\gamma)=\ln Z(\gamma)$

f is decreasing
f is convex
$f'(0) \geq -n$
$f(0) \leq \ln A$

Properties of partition functions

$$f(\gamma)=\ln Z(\gamma)$$

f is decreasing

f is convex

$f'(0) \geq -n$

$f(0) \leq \ln A$

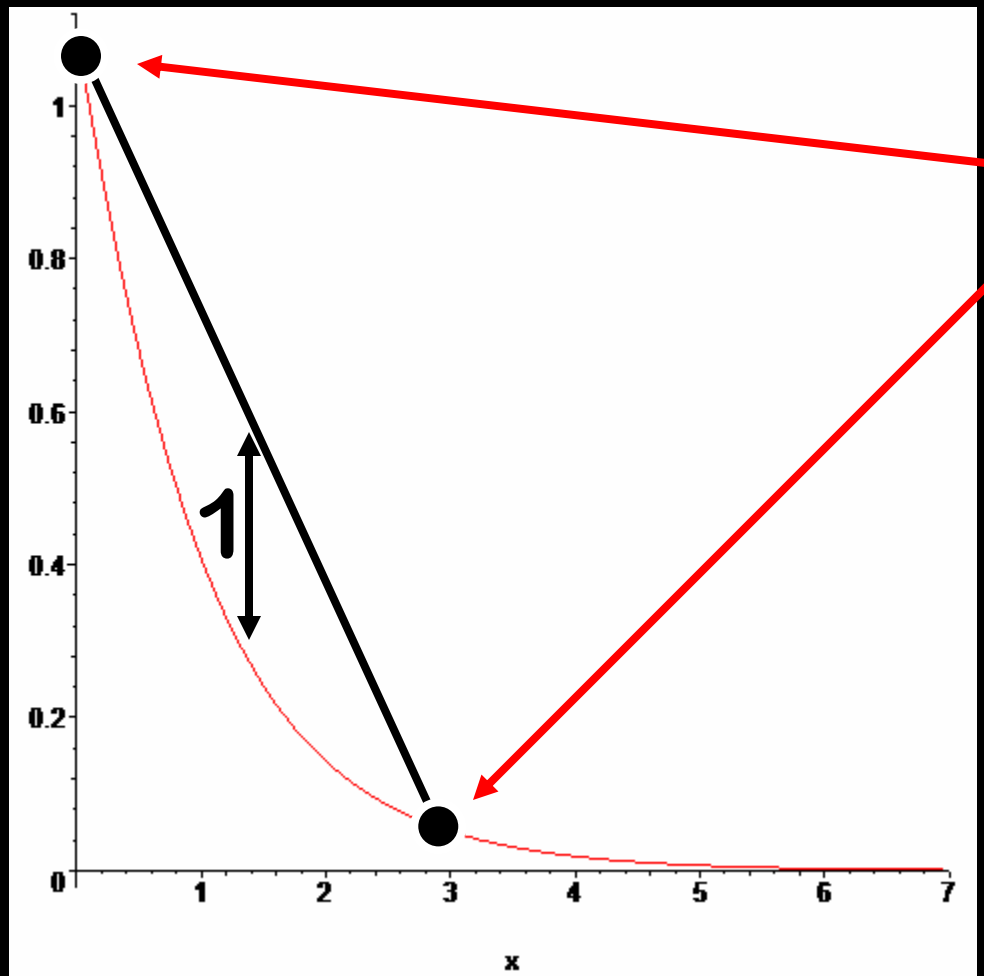
$$f(\beta) = \ln \sum_{k=0}^{n} a_k e^{-\beta k}$$

$$f'(\beta) = \frac{-\sum_{k=0}^{n} a_k k e^{-\beta k}}{\sum_{k=0}^{n} a_k e^{-\beta k}}$$

$$(\ln f)' = \frac{f'}{f}$$

GOAL: proving Lemma:
for every partition function there exists a cooling schedule of length $O^*((\ln A)^{1/2})$

$$f(\gamma)=\ln Z(\gamma)$$



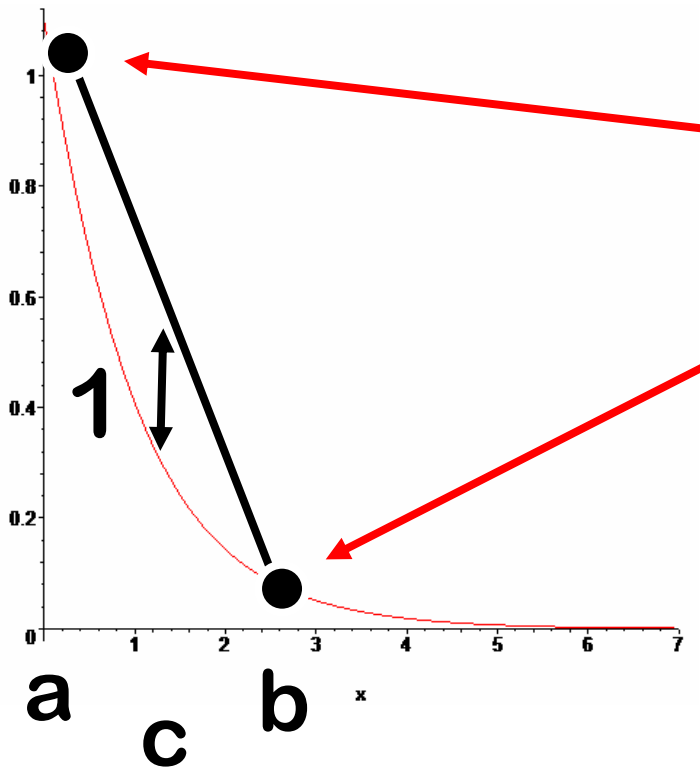
f is decreasing
f is convex
$f'(0) \geq -n$
$f(0) \leq \ln A$

either f or f' changes a lot

Let $K:=\Delta f$

Then

$$\Delta(\ln |f'|) \geq \frac{1}{K}$$



Let $K := \Delta f$

Then

$$\Delta(\ln |f'|) \geq \frac{1}{K}$$

$c := (a+b)/2$, $\Delta := b-a$

have $f(c) = (f(a)+f(b))/2 - 1$

f is convex ⇒

$(f(a) - f(c)) / \Delta \leq f'(a)$

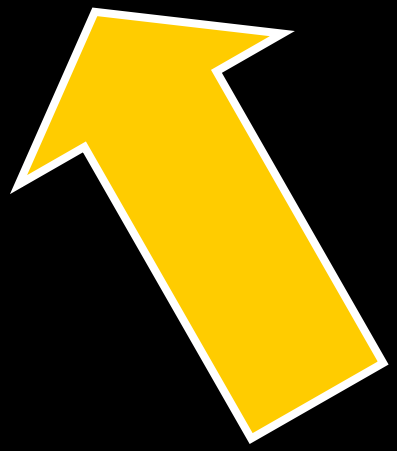
$(f(c) - f(b)) / \Delta \geq f'(b)$

$$\frac{f'(b)}{f'(a)} \leq 1 - 1/\Delta f \leq e^{-\Delta f}$$

Let $K := \Delta f$

Then

$$\Delta(\ln |f'|) \geq \frac{1}{K}$$



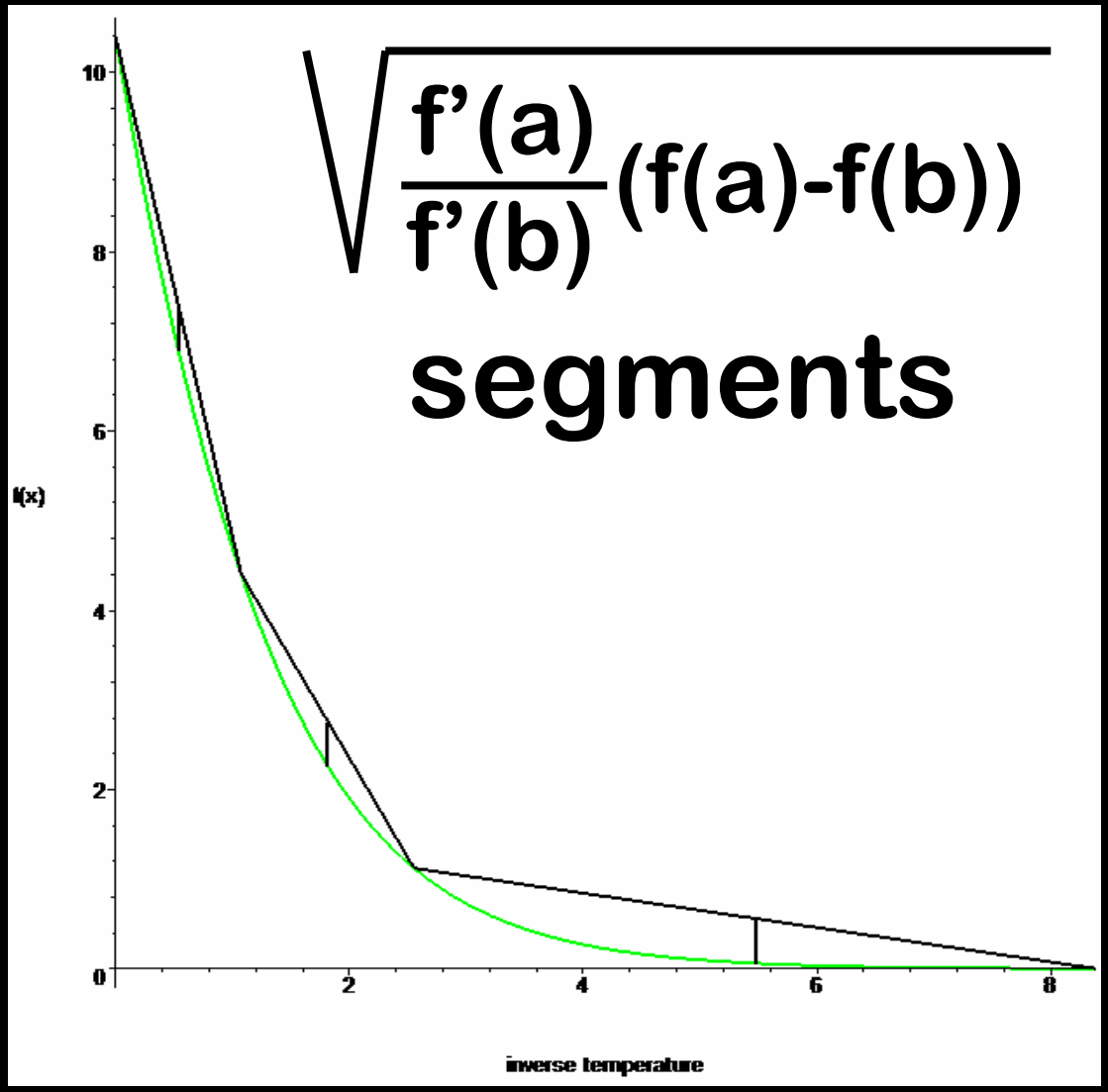
$c := (a+b)/2$, $\Delta := b-a$

have $f(c) = (f(a)+f(b))/2 - 1$

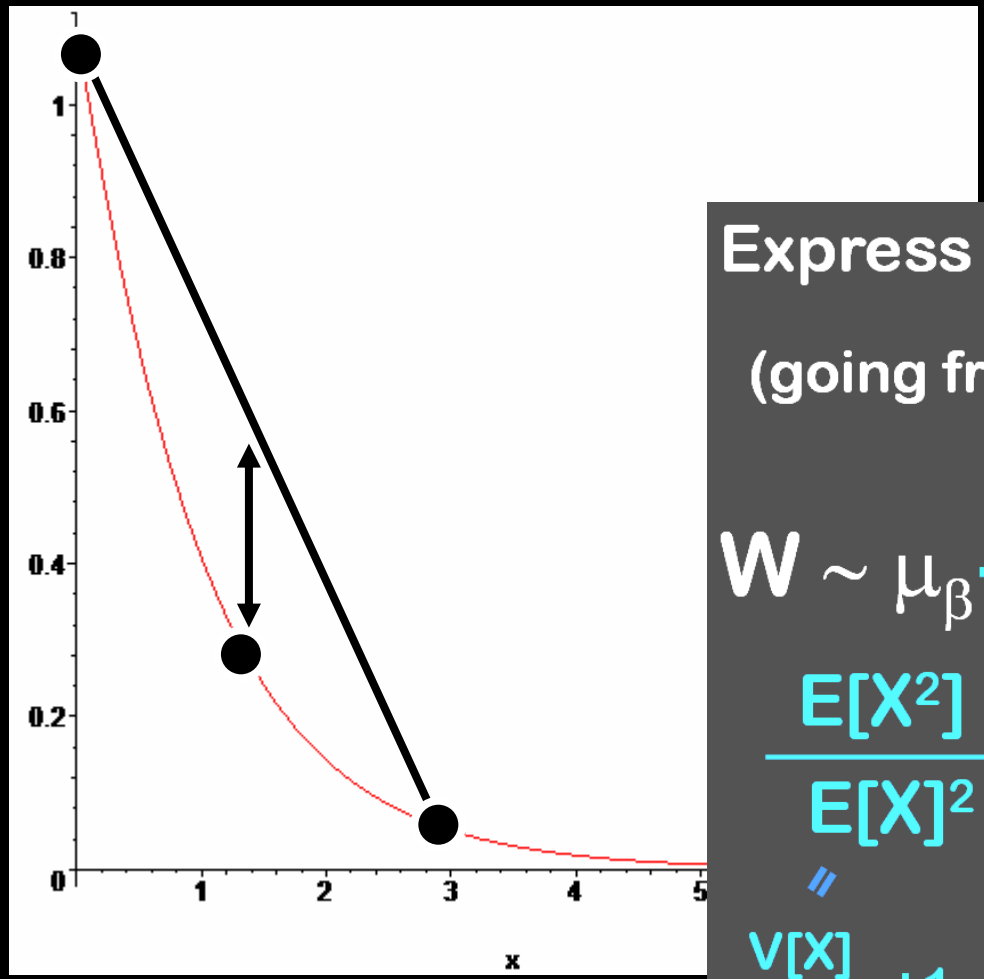
f is convex ➡

$$(f(a) - f(c)) / \Delta \leq f'(a)$$
$$(f(c) - f(b)) / \Delta \geq f'(b)$$

f:[a,b] → R, convex, decreasing can be "approximated" using



Technicality: getting to $2\alpha-\beta$



β α $2\alpha-\beta$

Express SCV using partition function

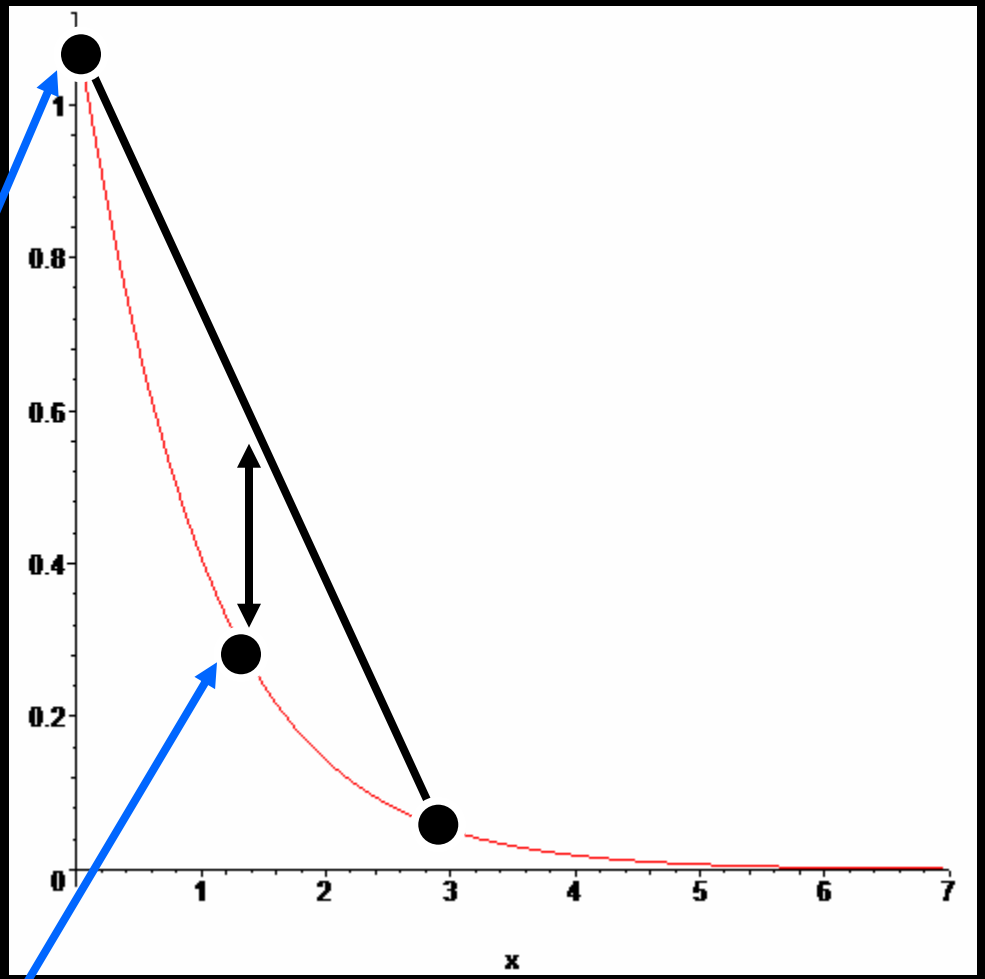
(going from β to α) $E[X] = \frac{Z(\alpha)}{Z(\beta)}$

$W \sim \mu_\beta \rightarrow X = \exp(H(W)(\beta - \alpha))$

$$\frac{E[X^2]}{E[X]^2} = \frac{Z(2\alpha-\beta)\,Z(\beta)}{Z(\alpha)^2} \leq c$$

$$\frac{V[X]}{E[X]^2} + 1$$

Technicality: getting to 2α-β

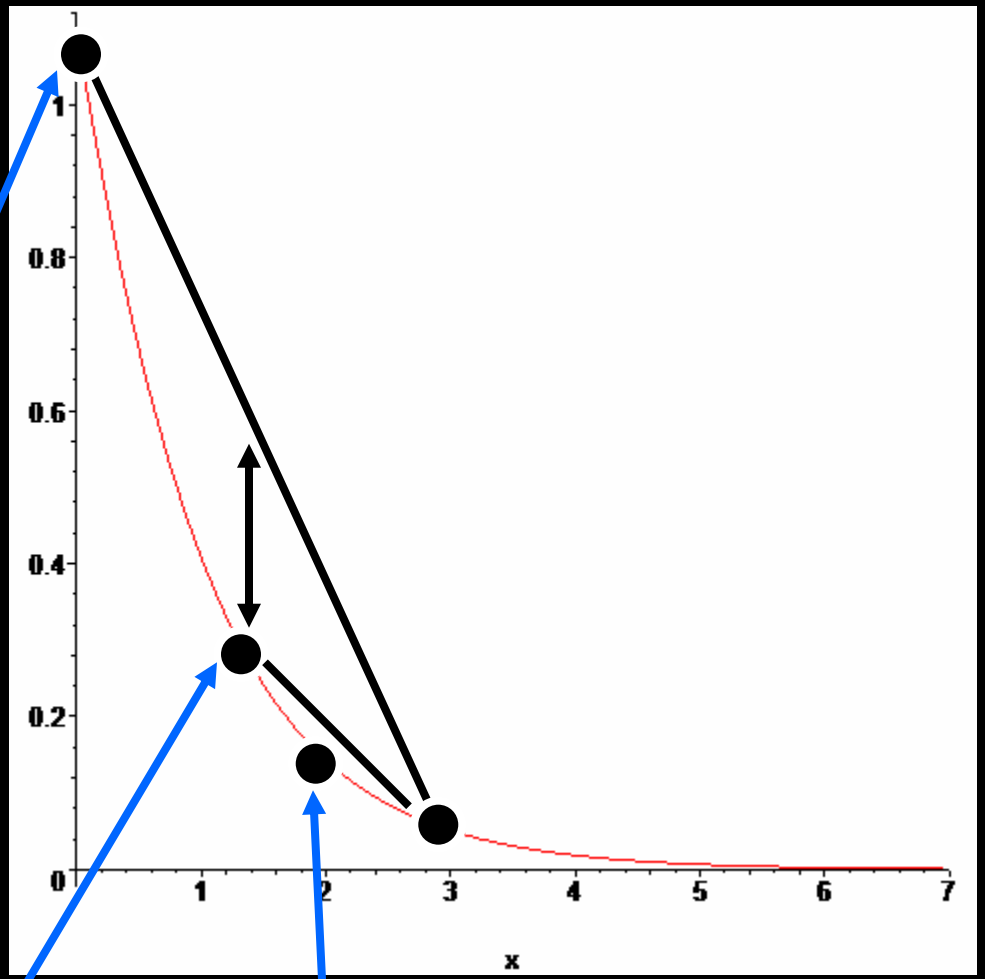


β_i

β_{i+1}

β α $2\alpha-\beta$

Technicality: getting to 2α-β



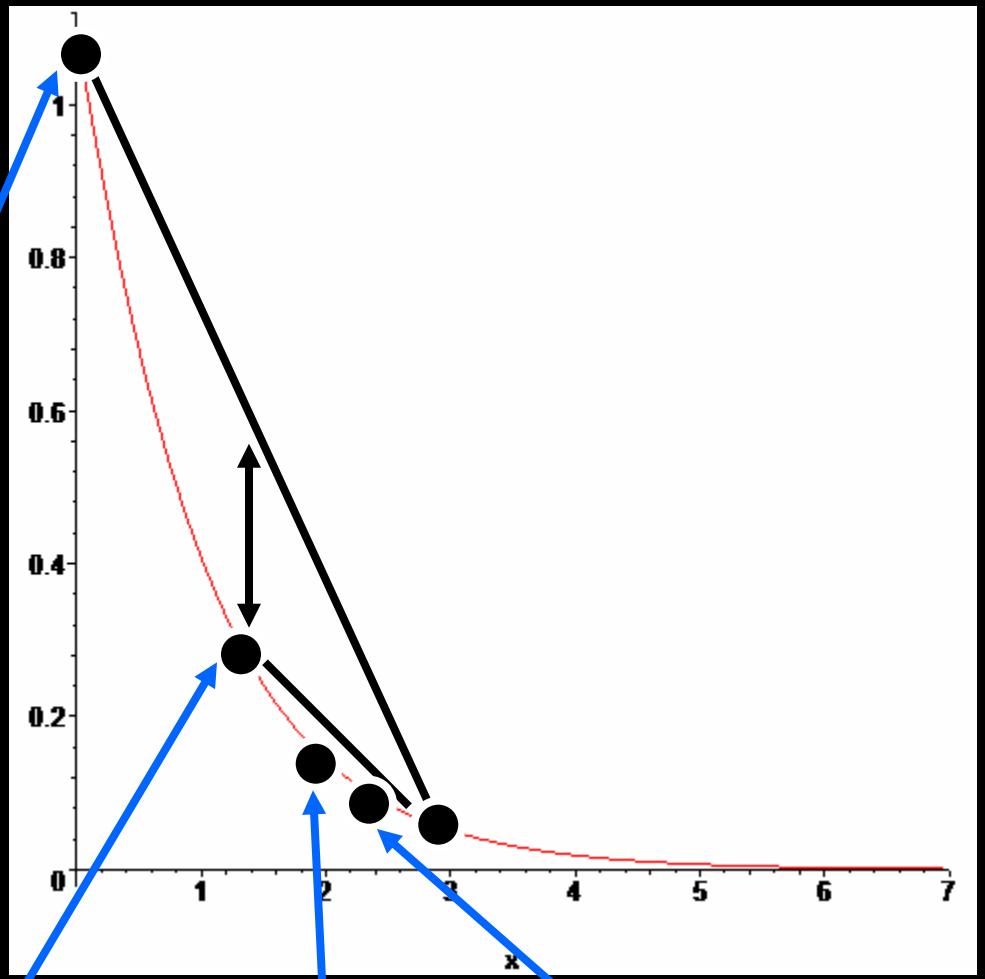
β_i

β α $2\alpha-\beta$

β_{i+1}

β_{i+2}

Technicality: getting to 2α-β



β_i

β α $2\alpha-\beta$

β_{i+1} β_{i+2} β_{i+3}

In ln A

extra steps

Existential → Algorithmic

~~there exists~~

can get adaptive schedule of length $O^*((\ln A)^{1/2})$

↓

can get adaptive schedule of length $O^*((\ln A)^{1/2})$

Algorithmic construction

Our main result:

using a sampler oracle for μ_β

$$\mu_\beta(x) = \frac{\exp(-H(x)\beta)}{Z(\beta)}$$

we can construct a cooling schedule of length

$$\leq 38\,(\ln A)^{1/2}(\ln \ln A)(\ln n)$$

Total number of oracle calls

$$\leq 10^7\,(\ln A)\,(\ln \ln A + \ln n)^7 \ln(1/\delta)$$

Algorithmic construction

current inverse temperature β

ideally move to α such that

$$B_1 \leq \frac{E[X^2]}{E[X]^2} \leq B_2$$

$$E[X] = \frac{Z(\alpha)}{Z(\beta)}$$

Algorithmic construction

current inverse temperature β

ideally move to α such that

$$B_1 \leq \frac{E[X^2]}{E[X]^2} \leq B_2$$

$$E[X] = \frac{Z(\alpha)}{Z(\beta)}$$

X is “easy to estimate”

Algorithmic construction

current inverse temperature β

ideally move to α such that

$$B_1 \leq \frac{E[X^2]}{E[X]^2} \leq B_2$$

$$E[X] = \frac{Z(\alpha)}{Z(\beta)}$$

we make progress (where $B_1 > 1$)

Algorithmic construction

current inverse temperature β

ideally move to α such that

$$B_1 \leq \frac{E[X^2]}{E[X]^2} \leq B_2$$

$$E[X] = \frac{Z(\alpha)}{Z(\beta)}$$

need to construct a “feeler” for this

Algorithmic construction

current inverse temperature β

ideally move to α such that

$$B_1 \leq \frac{E[X^2]}{E[X]^2} \leq B_2 \qquad E[X] = \frac{Z(\alpha)}{Z(\beta)}$$

$$\leqq \frac{Z(\beta)}{Z(\alpha)} \frac{Z(2\beta-\alpha)}{Z(\alpha)}$$

need to construct a “feeler” for this

Algorithmic construction

current inverse temperature β

ideally move to α such that

bad "feeler"

$$B_1 \leq \frac{E[X^2]}{E[X]^2} \leq B_2$$

$$E[X] = \frac{Z(\alpha)}{Z(\beta)}$$

$$\leqslant \frac{Z(\beta)}{Z(\alpha)} \frac{Z(2\beta-\alpha)}{Z(\alpha)}$$

need to construct a "feeler" for this

estimator for $\frac{Z(\beta)}{Z(\alpha)}$

$$Z(\beta) = \sum_{k=0}^{n} a_k \, e^{-\beta k}$$

For $W \sim \mu_\beta$ we have $P(H(W)=k) = \frac{a_k \, e^{-\beta k}}{Z(\beta)}$

estimator for $\frac{Z(\beta)}{Z(\alpha)}$

If H(X)=k likely at both α, $\beta \Rightarrow$ estimator

$$Z(\beta) = \sum_{k=0}^{n} a_k \, e^{-\beta k}$$

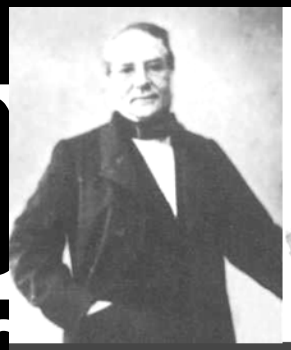
For $W \sim \mu_\beta$ we have $P(H(W)=k) = \frac{a_k \, e^{-\beta k}}{Z(\beta)}$

For $U \sim \mu_\alpha$ we have $P(H(U)=k) = \frac{a_k \, e^{-\alpha k}}{Z(\alpha)}$

estimator for $\frac{Z(\beta)}{Z(\alpha)}$

If H(X)=k likely at both α, $\beta \Rightarrow$ estimator

$$Z(\beta) = \sum_{k=0}^{n}$$



$$n = \frac{V[X]}{E[X]^2} \frac{1}{\varepsilon^2} \frac{1}{\delta}$$

Number of samples to achieve precision ε with confidence δ.

For W ~ μ_β w

For U ~ μ_α w

$$n = \frac{1}{E[X]} \frac{3}{\varepsilon^2} \ln(1/\delta)$$

$0 \le X \le 1$

estimator for $\frac{Z(\beta)}{Z(\alpha)}$

For $W \sim \mu_\beta$ we have $P(H(W)=k) = \frac{a_k e^{-\beta k}}{Z(\beta)}$

For $U \sim \mu_\alpha$ we have $P(H(U)=k) = \frac{a_k e^{-\alpha k}}{Z(\alpha)}$

$$\frac{P(H(U)=k)}{P(H(W)=k)} e^{k(\alpha-\beta)} = \frac{Z(\beta)}{Z(\alpha)}$$

estimator for $\frac{Z(\beta)}{Z(\alpha)}$

For $W \sim \mu_\beta$ we have $P(H(W)=k) = \frac{a_k e^{-\beta k}}{Z(\beta)}$

For $U \sim \mu_\alpha$ we have $P(H(U)=k) = \frac{a_k e^{-\alpha k}}{Z(\alpha)}$

$$\frac{P(H(U)=k)}{P(H(W)=k)} e^{k(\alpha-\beta)} = \frac{Z(\beta)}{Z(\alpha)}$$

PROBLEM: P(H(W)=k) can be too small

Rough estimator for $\frac{Z(\beta)}{Z(\alpha)}$

$$Z(\beta) = \sum_{k=0}^{n} a_k \, e^{-\beta k}$$

interval instead of single value

For $W \sim \mu_\beta$ we have

$$P(H(W) \in [c,d]) = \frac{\sum_{k=c}^{d} a_k \, e^{-\beta k}}{Z(\beta)}$$

For $U \sim \mu_\alpha$ we have

$$P(H(W) \in [c,d]) = \frac{\sum_{k=c}^{d} a_k \, e^{-\alpha k}}{Z(\alpha)}$$

Rough estimator for $\frac{Z(\beta)}{Z(\alpha)}$

If $|\alpha-\beta| \cdot |d-c| \leq 1$ then

$$\frac{1}{e}\,\frac{Z(\beta)}{Z(\alpha)} \leq \frac{P(H(U)\in[c,d])}{P(H(W)\in[c,d])}\, e^{c(\alpha-\beta)} \leq e\,\frac{Z(\beta)}{Z(\alpha)}$$

We also need $P(H(U) \in [c,d])$
$P(H(W) \in [c,d])$ to be large.

$$\frac{\sum_{k=c}^{d} a_k\, e^{-\alpha k}}{\sum_{k=c}^{d} a_k\, e^{-\beta k}}\, e^{c(\alpha-\beta)} = \frac{\sum_{k=c}^{d} a_k\, e^{-\alpha (k-c)}}{\sum_{k=c}^{d} a_k\, e^{-\beta (k-c)}}$$

We will:

Split {0,1,…,n} into $h \le 4(\ln n)\sqrt{\ln A}$ intervals

[0],[1],[2],…,[c,c(1+1/$\sqrt{\ln A}$)],…

for any inverse temperature β there exists a interval with $P(H(W) \in I) \ge 1/8h$

We say that I is **HEAVY** for β

We will:

Split {0,1,…,n} in
intervals
[0],[1],[2],…,[

Rough estimator for $\frac{Z(\beta)}{Z(\alpha)}$

If $|\alpha-\beta| \cdot |d-c| \le 1$ then

$$\frac{1}{e}\frac{Z(\beta)}{Z(\alpha)} \le \frac{P(H(U)\in[c,d])}{P(H(W)\in[c,d])} e^{c(\alpha-\beta)} \le e\frac{Z(\beta)}{Z(\alpha)}$$

We also need $P(H(U) \in [c,d])$
$P(H(W) \in [c,d])$ to be large.

$$\frac{\sum_{k=c}^{d} a_k e^{-\alpha k}}{\sum_{k=c}^{d} a_k e^{-\beta k}} e^{c(\alpha-\beta)} = \frac{\sum_{k=c}^{d} a_k e^{-\alpha (k-c)}}{\sum_{k=c}^{d} a_k e^{-\beta (k-c)}}$$

for any inverse temperature β there exists a interval with $P(H(W)\in I) \ge 1/8h$

We say that I is **HEAVY** for β

Algorithm

repeat

find an interval I which is heavy for
the current inverse temperature β

see how far I is heavy (until some β^*)

use the interval I for the feeler $\frac{Z(\beta)}{Z(\alpha)}$ $\frac{Z(2\beta-\alpha)}{Z(\alpha)}$

ANALYSIS:

either

* make progress, or
* eliminate the interval I
* or make a "long move"

Algorithm

repeat

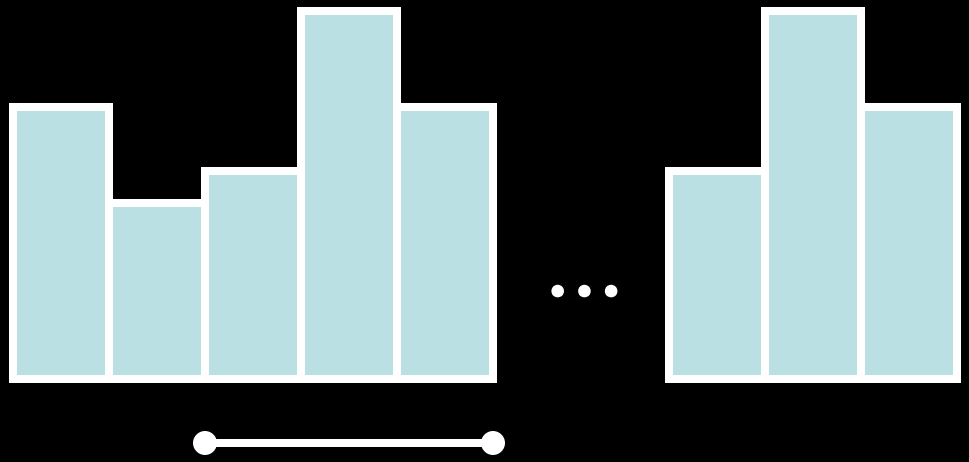
find an interval I which is heavy for the current inverse temperature β

see how far I is heavy (until some β^*)

use the interval I for the feeler $\frac{Z(\beta)}{Z(\alpha)}$ $\frac{Z(2\beta-\alpha)}{Z(\alpha)}$

I is heavy

distribution of h(X) where $X \sim \mu_\beta$



I = a heavy interval at β

Algorithm

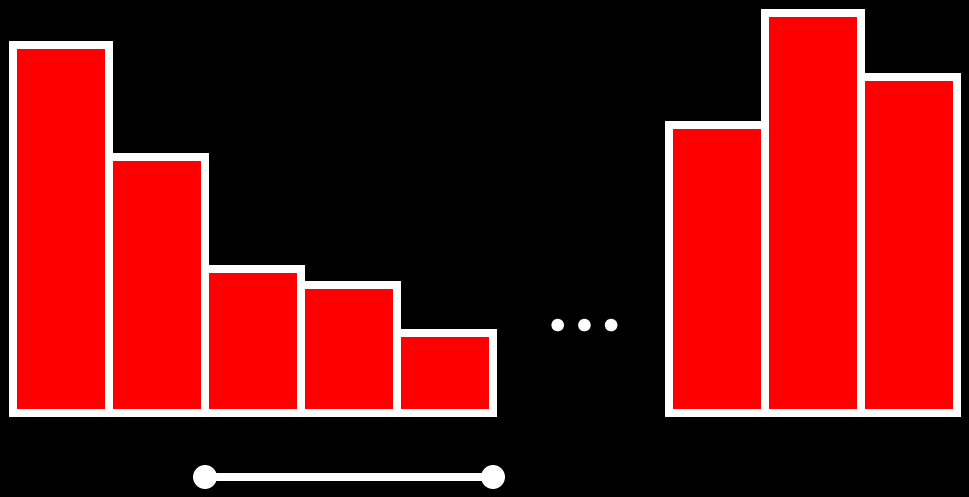
repeat

find an interval I which is heavy for the current inverse temperature β

see how far I is heavy (until some β^*)

use the interval I for the feeler $\frac{Z(\beta)}{Z(\alpha)}$ $\frac{Z(2\beta-\alpha)}{Z(\alpha)}$

distribution of h(X) where $X \sim \mu_\gamma$

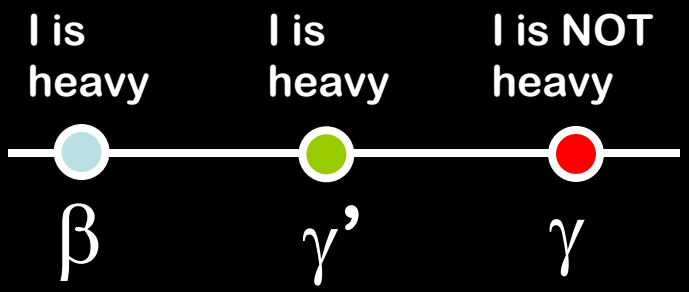


no longer heavy at γ !

I = a heavy interval at β

Algorithm

repeat

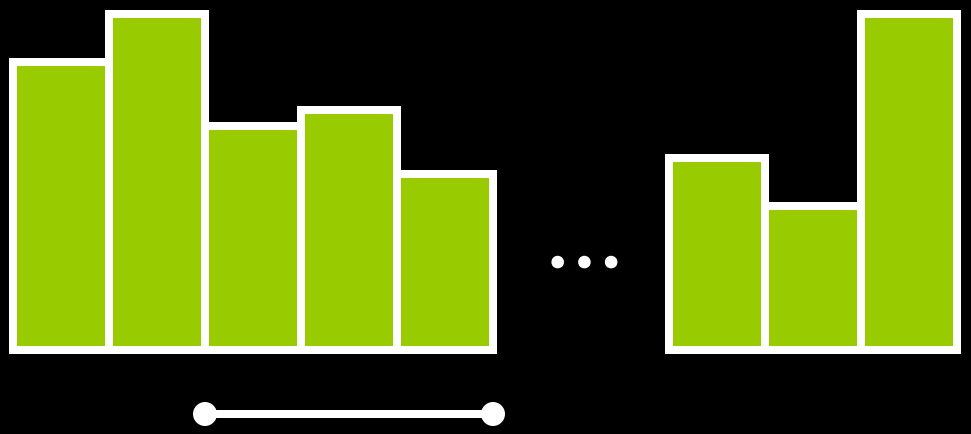


find an interval I which is heavy for the current inverse temperature β

see how far I is heavy (until some β^*)

use the interval I for the feeler $\frac{Z(\beta)}{Z(\alpha)}$ $\frac{Z(2\beta-\alpha)}{Z(\alpha)}$

distribution of h(X) where $X \sim \mu_{\gamma'}$



heavy at γ'

I = a heavy interval at β

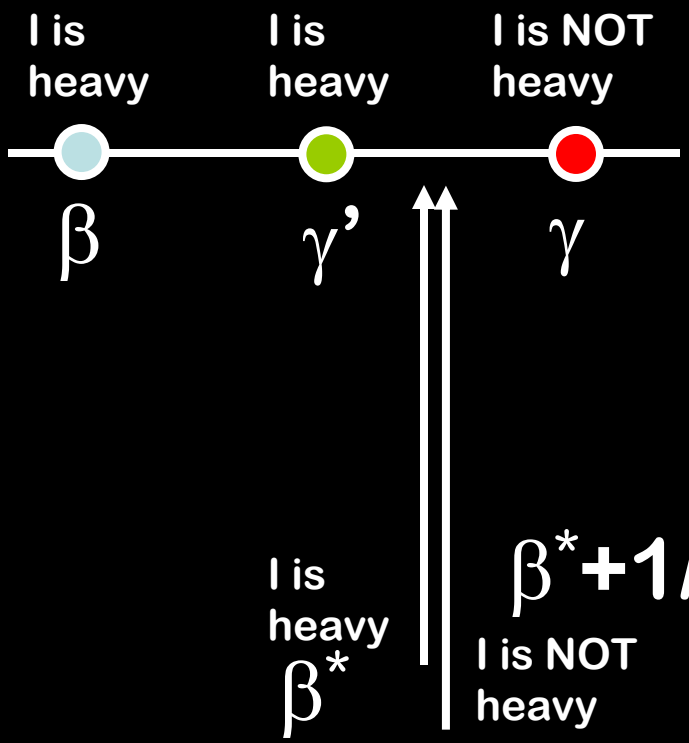
Algorithm

repeat

find an interval I which is heavy for the current inverse temperature β

see how far I is heavy (until some β^*)

use the interval I for the feeler $\frac{Z(\beta)}{Z(\alpha)}$ $\frac{Z(2\beta-\alpha)}{Z(\alpha)}$



I=[a,b]

use binary search to find β^*

α = min{1/(b-a), ln A}

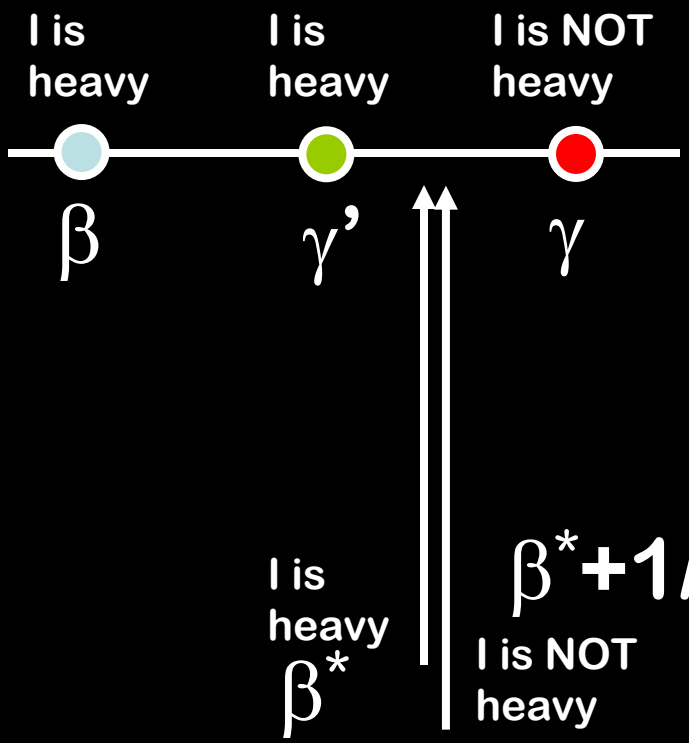
Algorithm

repeat

find an interval I which is heavy for the current inverse temperature β

see how far I is heavy (until some β^*)

use the interval I for the feeler $\frac{Z(\beta)}{Z(\alpha)}$ $\frac{Z(2\beta-\alpha)}{Z(\alpha)}$



I=[a,b]

use binary search to find β^*

α = min{1/(b-a), ln A}

How do you know that you can use binary search?

How do you know that you can use binary search?

Lemma: the set of temperatures for which I is h-heavy is an interval.

I is h-heavy at β $\Leftrightarrow$ $P(h(X) \in I) \geq 1/8h$ for $X \sim \mu_\beta$

$$\sum_{k \in I} a_k e^{-\beta k} \geq \frac{1}{8h} \sum_{k=0}^{n} a_k e^{-\beta k}$$

How do you know that you can use binary search?

$$\sum_{k \in I} a_k \, e^{-\beta k} \geq \frac{1}{8h} \sum_{k=0}^{n} a_k \, e^{-\beta k}$$

$x = e^{-\beta}$

Descarte's rule of signs:

$$c_0 x^0 + c_1 x^1 + c_2 x^2 + \ldots + c_n x^n$$

\+ \- \+ \+

sign change

number of <u>positive</u> roots $\leq$ number of sign changes

How do you know that you can use binary search?

$$-1+x+x^2+x^3+\ldots+x^n$$

$$1+x+x^{20}$$

Descarte's rule of signs:

$$c_0 x^0 + c_1 x^1 + c_2 x^2 + \ldots + c_n x^n$$

$$+ \quad - \quad + \qquad\qquad +$$

sign change

number of positive roots $\leq$ number of sign changes

How do you know that you can use binary search?

$$\sum_{k \in I} a_k \, e^{-\beta k} \geq \frac{1}{8h} \sum_{k=0}^{n} a_k \, e^{-\beta k}$$

$x = e^{-\beta}$

Descarte's rule of signs:

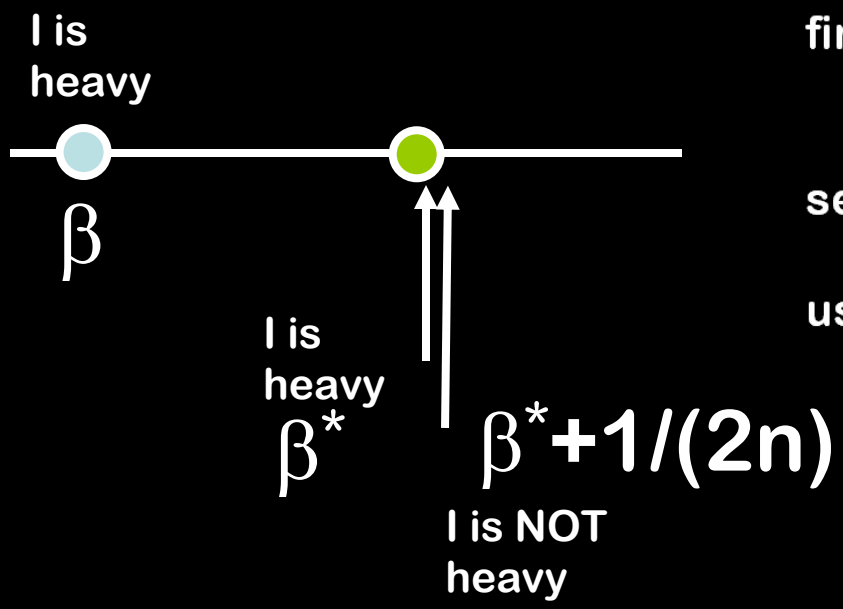
$$c_0 x^0 + c_1 x^1 + c_2 x^2 + \ldots + c_n x^n$$

\+ \- \+ \+

sign change

number of <u>positive</u> roots $\leq$ number of sign changes

I=[a,b]



Algorithm

repeat

find an interval I which is heavy for the current inverse temperature β

see how far I is heavy (until some β^*)

use the interval I for the feeler $\frac{Z(\beta)}{Z(\alpha)}$ $\frac{Z(2\beta-\alpha)}{Z(\alpha)}$

can roughly compute ratio of $Z(\alpha)/Z(\alpha')$ for $\alpha, \alpha' \in [\beta,\beta^*]$ if $|\alpha-\alpha|.|b-a| \le 1$

Rough estimator for $\frac{Z(\beta)}{Z(\alpha)}$

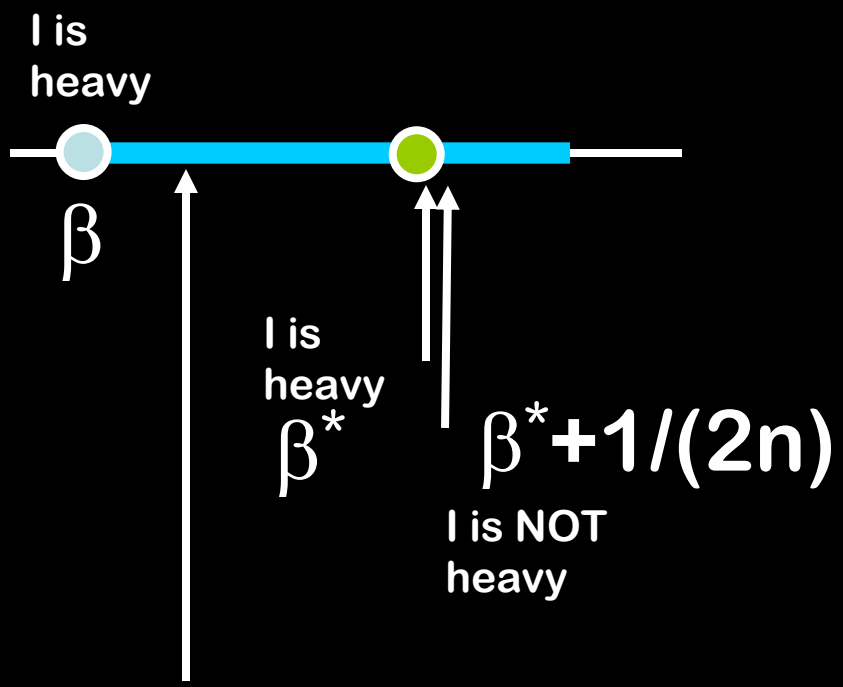
If $|\alpha-\beta| \cdot |d-c| \le 1$ then

$$\frac{1}{e}\frac{Z(\beta)}{Z(\alpha)} \le \frac{P(H(U)\in[c,d])}{P(H(W)\in[c,d])} e^{c(\alpha-\beta)} \le e\frac{Z(\beta)}{Z(\alpha)}$$

We also need $P(H(U) \in [c,d])$ $P(H(W) \in [c,d])$ to be large

$$\frac{\sum_{k=c}^{d} a_k e^{-\alpha k}}{\sum_{k=c}^{d} a_k e^{-\beta k}} e^{c(\alpha-\beta)} = \frac{\sum_{k=c}^{d} a_k e^{-\alpha (k-c)}}{\sum_{k=c}^{d} a_k e^{-\beta (k-c)}}$$

I=[a,b]



1. success

2. eliminate interval

3. long move

find largest α such that

$$\frac{Z(\beta)}{Z(\alpha)} \quad \frac{Z(2\beta-\alpha)}{Z(\alpha)} \leq C$$

can roughly compute ratio of $Z(\alpha)/Z(\alpha')$ for $\alpha, \alpha' \in [\beta,\beta^*]$ if $|\alpha-\alpha|.|b-a| \leq 1$

input : A black-box sampler for $X \sim \mu_\beta$ for any $\beta \geq 0$, starting inverse temperature β_0.
output: A cooling schedule for Z.

Bad $\leftarrow \emptyset$
print β_0
if $\beta_0 < \ln A$ **then**
  $I \leftarrow$FIND-HEAVY(β_0, Bad)
  $w \leftarrow$ the width of I
  $L \leftarrow \min\{\beta_0 + 1/w, \ln A\}$; (where $1/0 = \infty$)
  $\beta^* \leftarrow$ binary search on $\beta^* \in [\beta_0, L]$ with precision $1/(2n)$, using predicate IS-HEAVY(β^*, I)
  $\beta \leftarrow$ binary search on $\beta \in [\beta_0, (\beta^* + \beta_0)/2]$ with precision $1/(4n)$, using predicate $\text{EST}(I, \beta_0, \beta) \cdot \text{EST}(I, 2\beta - \beta_0, \beta) \leq 2000$
  if $\beta < (\beta^* + \beta_0)/2$ **then**
    PRINT-COOLING-SCHEDULE(β) (optimal move)
  else
    if $\beta = L$ **then**
      PRINT-COOLING-SCHEDULE(β) (long move)
    else
      $\gamma \leftarrow (\beta^* - \beta_0)/2$
      print $\beta_0 + \gamma, \beta_0 + (3/2)\gamma, \beta_0 + (7/4)\gamma, \ldots, \beta_0 + (2 - 2^{-\lceil \ln \ln A \rceil})\gamma$
      Bad $\leftarrow$ Bad $\cup I$
      PRINT-COOLING-SCHEDULE(β^*) (interval move)
    end
  end
else
  print ∞
end

if we have sampler oracles for μ_β then we can get adaptive schedule of length $t=O^*(\ (\ln A)^{1/2}\ )$

independent sets $O^*(n^2)$
(using Vigoda'01, Dyer-Greenhill'01)

matchings $O^*(n^2m)$
(using Jerrum, Sinclair'89)

spin systems:
Ising model $O^*(n^2)$ for $\beta<\beta_c$
(using Marinelli, Olivieri'95)
k-colorings $O^*(n^2)$ for $k>2\Delta$
(using Jerrum'95)

Outline

1. Counting problems

2. Basic tools: Chernoff, Chebyshev

3. Dealing with large quantities
   (the product method)

4. Statistical physics

5. Cooling schedules (our work)

6. More...

Outline

6. More...

   a) proof of Dyer-Frieze

   b) independent sets revisited

   c) warm starts

Appendix – proof of:

1) $E[X_1 X_2 \dots X_t]$ = "WANTED"

2) the X_i are easy to estimate

$$\frac{V[X_i]}{E[X_i]^2} = O(1)$$

Theorem (Dyer-Frieze'91)

$O(t^2/\varepsilon^2)$ samples ($O(t/\varepsilon^2)$ from each X_i)

give

$1\pm\varepsilon$ estimator of "WANTED" with prob$\geq$3/4

How precise do the X_i have to be?

First attempt – term by term

Main idea:

$$(1\pm\frac{\varepsilon}{t})(1\pm\frac{\varepsilon}{t})(1\pm\frac{\varepsilon}{t})\ldots(1\pm\frac{\varepsilon}{t}) \approx 1\pm\varepsilon$$

$$n = \Theta\left(\frac{V[X]}{E[X]^2}\frac{1}{\varepsilon^2}\ln(1/\delta)\right)$$

each term $\Omega(t^2)$ samples $\Rightarrow$ $\Omega(t^3)$ total

How precise do the X_i have to be?

Analyzing SCV is better (Dyer-Frieze'1991)

$X = X_1 X_2 \dots X_t$

GOAL: SCV(X) $\leq \varepsilon^2/4$

squared coefficient of variation (SCV)

P(X gives $(1 \pm \varepsilon)$-estimate)

$$\geq 1 - \frac{V[X]}{E[X]^2} \frac{1}{\varepsilon^2}$$

How precise do the X_i have to be?

Analyzing SCV is better (Dyer-Frieze'1991)

Main idea:

$$SCV(X_i) \le \frac{\varepsilon^2/4}{t} \Rightarrow SCV(X) \lessapprox \varepsilon^2/4$$

proof:

$$SCV(X) = (1+SCV(X_1)) \dots (1+SCV(X_t)) - 1$$

$$SCV(X) = \frac{V[X]}{E[X]^2} = \frac{E[X^2]}{E[X]^2} - 1$$

How pre

Analy
(Dy

Main idea:

X_1, X_2 independent $\Rightarrow E[X_1 X_2] = E[X_1]E[X_2]$

X_1, X_2 independent $\Rightarrow X_1^2, X_2^2$ independent

X_1, X_2 independent $\Rightarrow$
$SCV(X_1X_2)=(1+SCV(X_1))(1+SCV(X_2))-1$

$$SCV(X_i) \le \frac{\varepsilon^2/4}{t} \Rightarrow SCV(X) \lessapprox \varepsilon^2/4$$

proof:

$$SCV(X) = (1+SCV(X_1)) \dots (1+SCV(X_t)) - 1$$

$$SCV(X)=\frac{V[X]}{E[X]^2} = \frac{E[X^2]}{E[X]^2} -1$$

How precise do the X_i have to be?

Analyzing SCV is better (Dyer-Frieze'1991)

$$X = X_1 \; X_2 \ldots X_t$$

Main idea:

$$SCV(X_i) \leq \frac{\varepsilon^2/4}{t} \Rightarrow SCV(X) \lessapprox \varepsilon^2/4$$

each term $O(t/\varepsilon^2)$ samples $\Rightarrow$ $O(t^2/\varepsilon^2)$ total

Outline

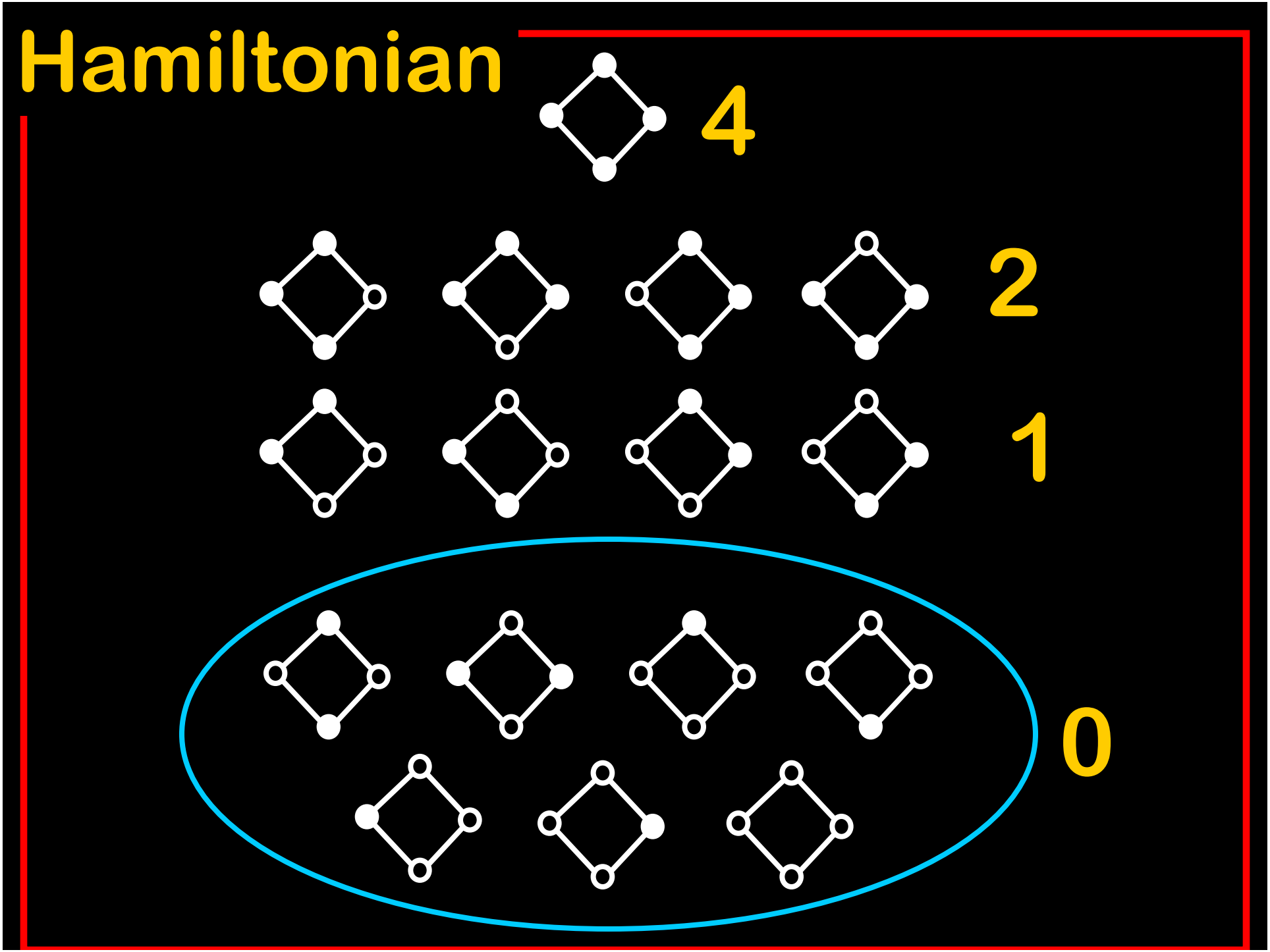
6. More...

a) proof of Dyer-Frieze

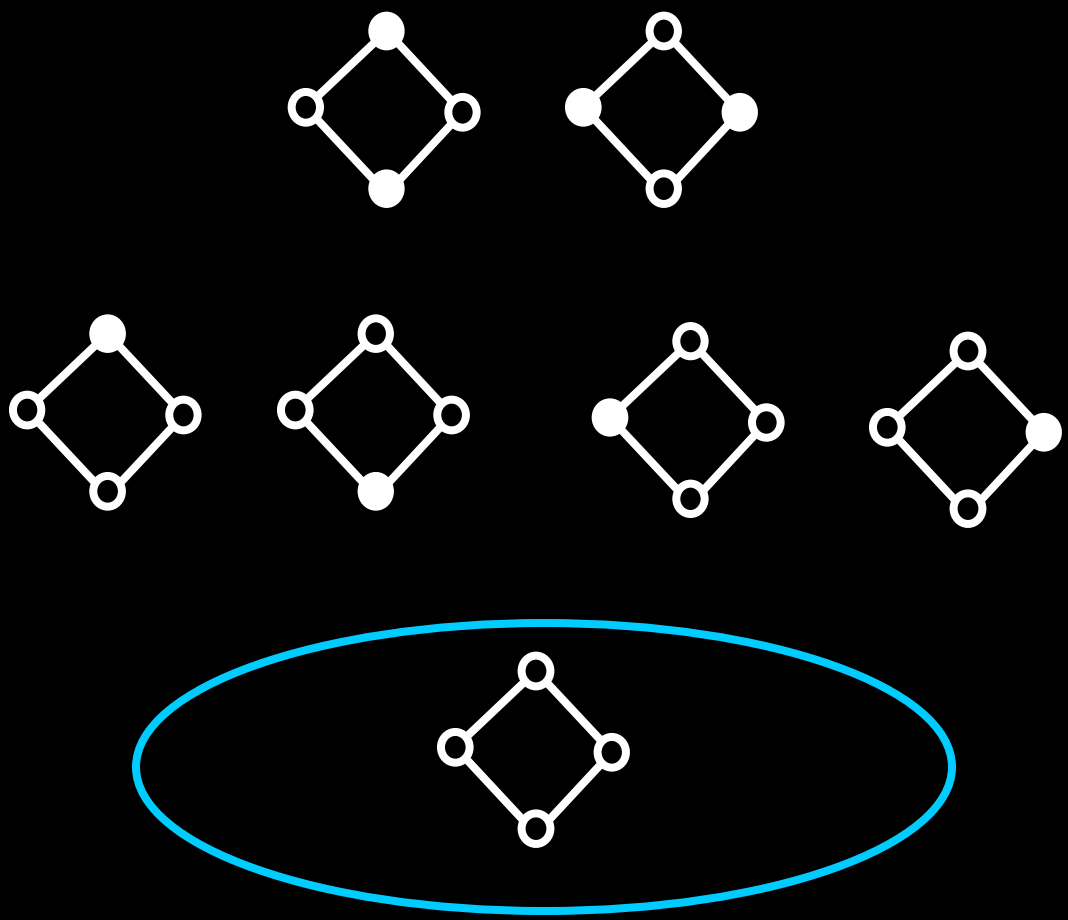
b) independent sets revisited

c) warm starts

Hamiltonian



Hamiltonian – many possibilities



(hardcore lattice gas model)

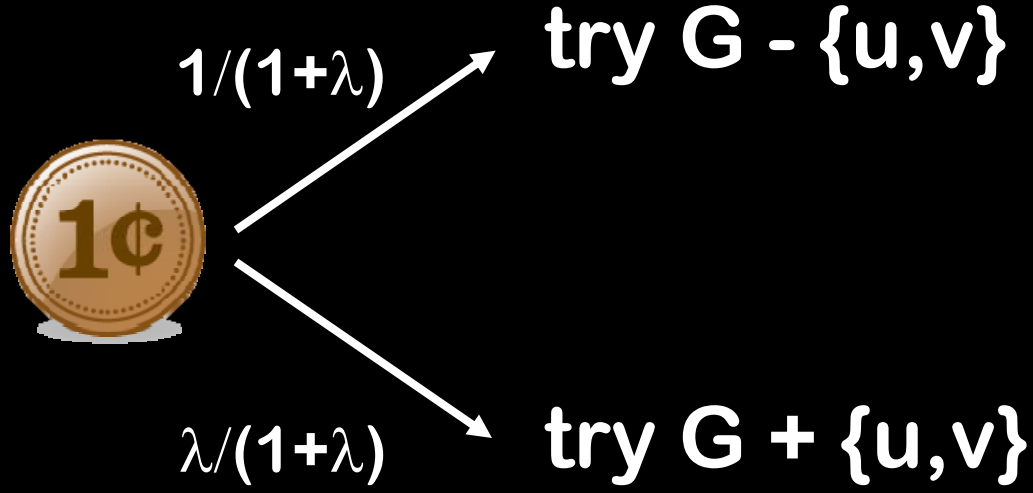
What would be a natural hamiltonian for planar graphs?

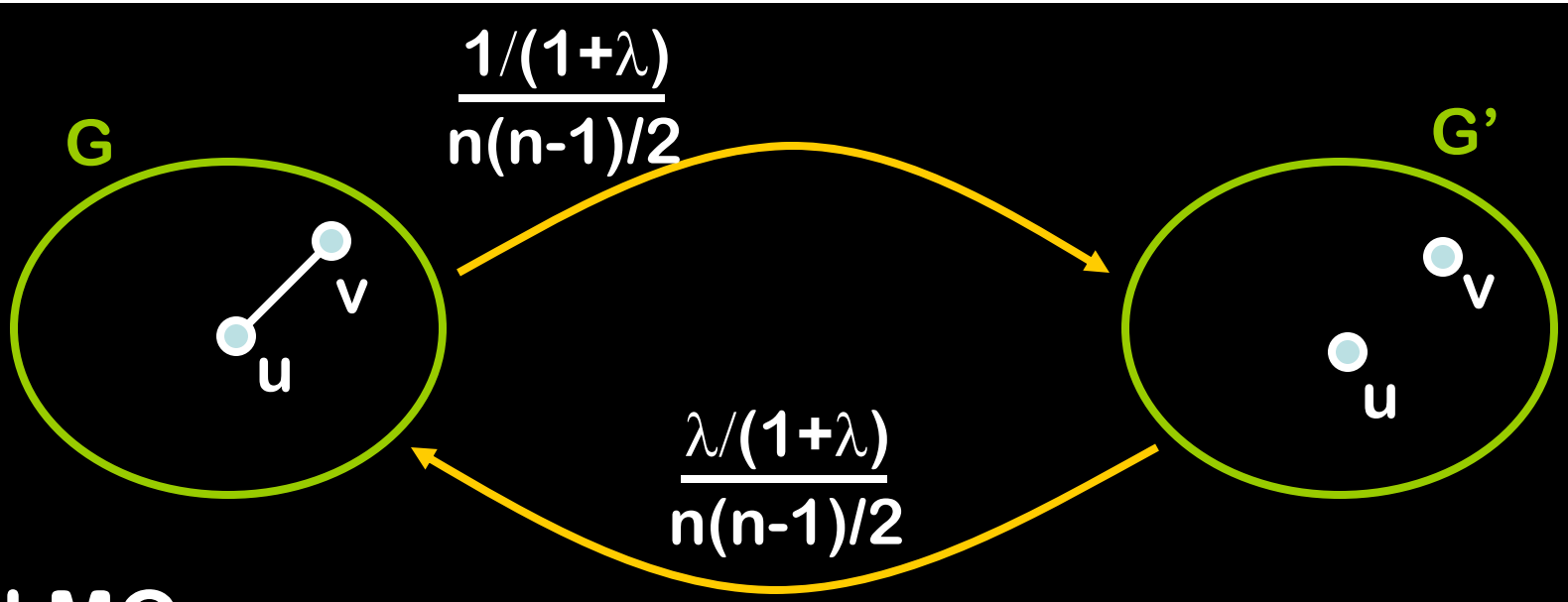
What would be a natural hamiltonian for planar graphs?

H(G) = number of edges

natural MC

pick u,v uniformly at random



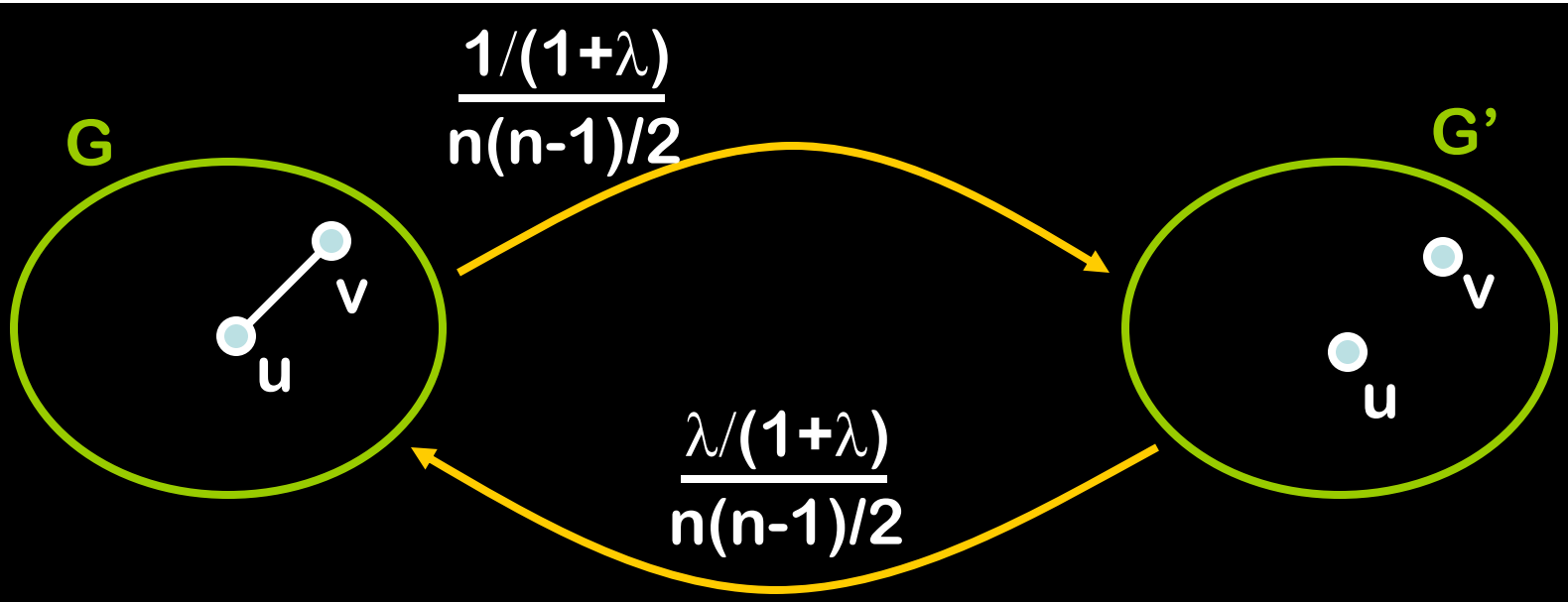


natural MC

pick u,v uniformly at random

1/(1+λ) try G - {u,v}

λ/(1+λ) try G + {u,v}



$\pi(\mathbf{G}) \propto \lambda^{\text{number of edges}}$

satisfies the detailed balance condition

$$\pi(G)\, P(G,G') = \pi(G')\, P(G',G)$$

($\lambda = \exp(-\beta)$)

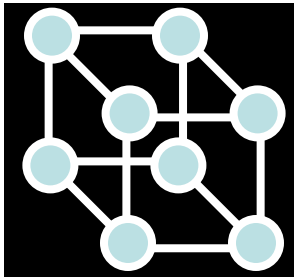
Outline

6. More...

a) proof of Dyer-Frieze

b) independent sets revisited

c) warm starts



(n=3)

Mixing time:

τ_{mix} = smallest t such that $| \mu_t - \pi |_{TV} \leq 1/e$

$\Theta(n \ln n)$

Relaxation time:

$\tau_{rel} = 1/(1-\lambda_2)$

$\Theta(n)$

$$\tau_{rel} \leq \tau_{mix} \leq \tau_{rel} \ln (1/\pi_{min})$$

(discrepancy may be substantially bigger for, e.g., matchings)

Estimating $\pi(S)$

Mixing time:

τ_{mix} = smallest t such that

$| \mu_t - \pi |_{TV} \leq 1/e$

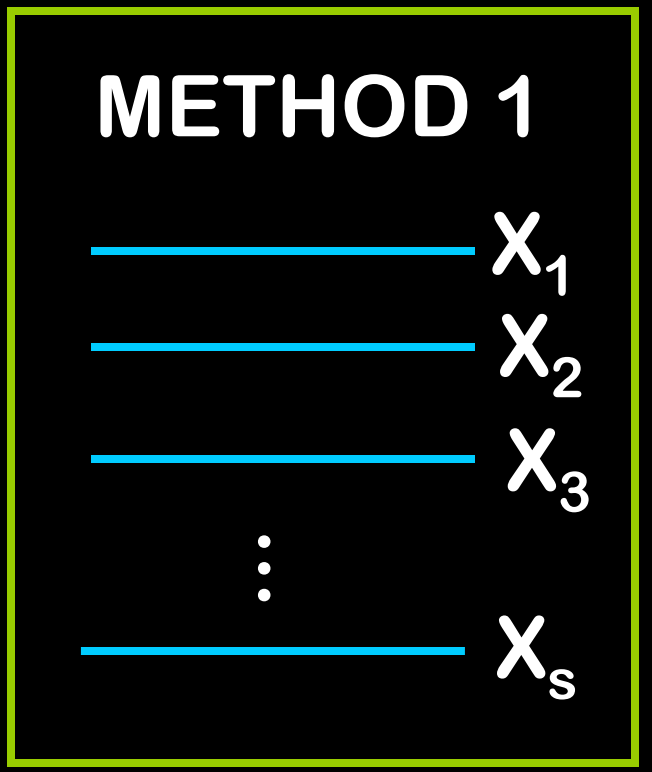
Relaxation time:

$\tau_{rel} = 1/(1-\lambda_2)$

$X \sim \pi$

$$Y = \begin{cases} 1 \text{ if } X \in S \\ 0 \text{ otherwise} \end{cases}$$

$E[Y] = \pi(S)$



Estimating $\pi(S)$

Mixing time:

τ_{mix} = smallest t such that

$\| \mu_t - \pi \|_{TV} \leq 1/e$

Relaxation time:

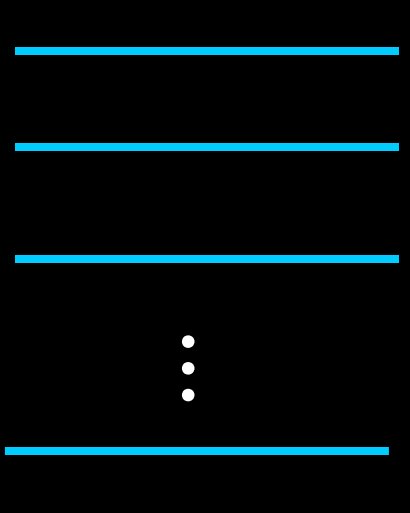
$\tau_{rel} = 1/(1-\lambda_2)$

$X \sim \pi$

$$Y = \begin{cases} 1 \text{ if } X \in S \\ 0 \text{ otherwise} \end{cases}$$

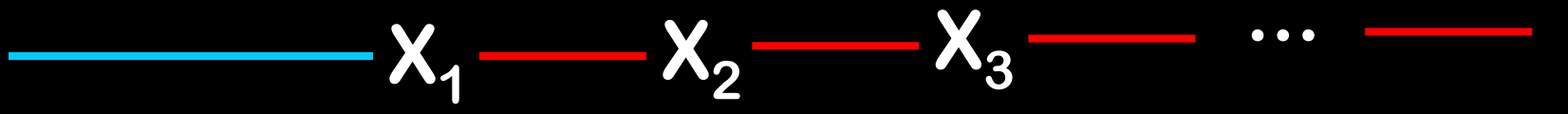
$E[Y] = \pi(S)$

METHOD 1



METHOD 2

(Gillman'98, Kahale'96, ...)



Further speed-up

Mixing time:

τ_{mix} = smallest t such that

$| \mu_t - \pi |_{TV} \leq 1/e$

Relaxation time:

$\tau_{rel} = 1/(1-\lambda_2)$

$$|\mu_t - \pi|_{TV} \leq \exp(-t/\tau_{rel})\ \mathrm{Var}_\pi(\mu_0/\pi)$$

$$\left(\sum \pi(x)(\mu_0(x)/\pi(x)-1)^2\right)^{1/2}$$

small $\Rightarrow$ called warm start

METHOD 2

(Gillman'98, Kahale'96, ...)

$X_1 \text{ —— } X_2 \text{ —— } X_3 \text{ —— } \cdots \text{ —— } X_s$

Further speed-up

sample at β can be used as a warm start for β'

$\Leftrightarrow$

cooling schedule can step from β' to β

$$\|\mu_t - \pi\|_{TV} \leq \exp(-t/\tau_{rel})\ \mathbf{Var}_\pi(\mu_0/\pi)$$

$$\left(\sum \pi(x)(\mu_0(x)/\pi(x)-1)^2\right)^{1/2}$$

small $\Rightarrow$ called warm start

METHOD 2

(Gillman'98, Kahale'96, ...)

$X_1 \quad X_2 \quad X_3 \quad \cdots \quad X_s$

sample at β can be used as a warm start for β'

$\Leftrightarrow$

cooling schedule can step from β' to β

revious cooling schedules

$Z(0) = A$ $\quad H:\Omega \to \{0,\ldots,n\}$

$\beta_0 = 0 < \beta_1 < \beta_2 < \ldots < \beta_t = \infty$

"Safe steps"

$\beta \to \beta + 1/n$

$\beta \to \beta\,(1 + 1/\ln A)$

$\ln A \to \infty$

(Bezáková,Štefankovič, Vigoda,V.Vazirani'06)

Cooling schedules of length

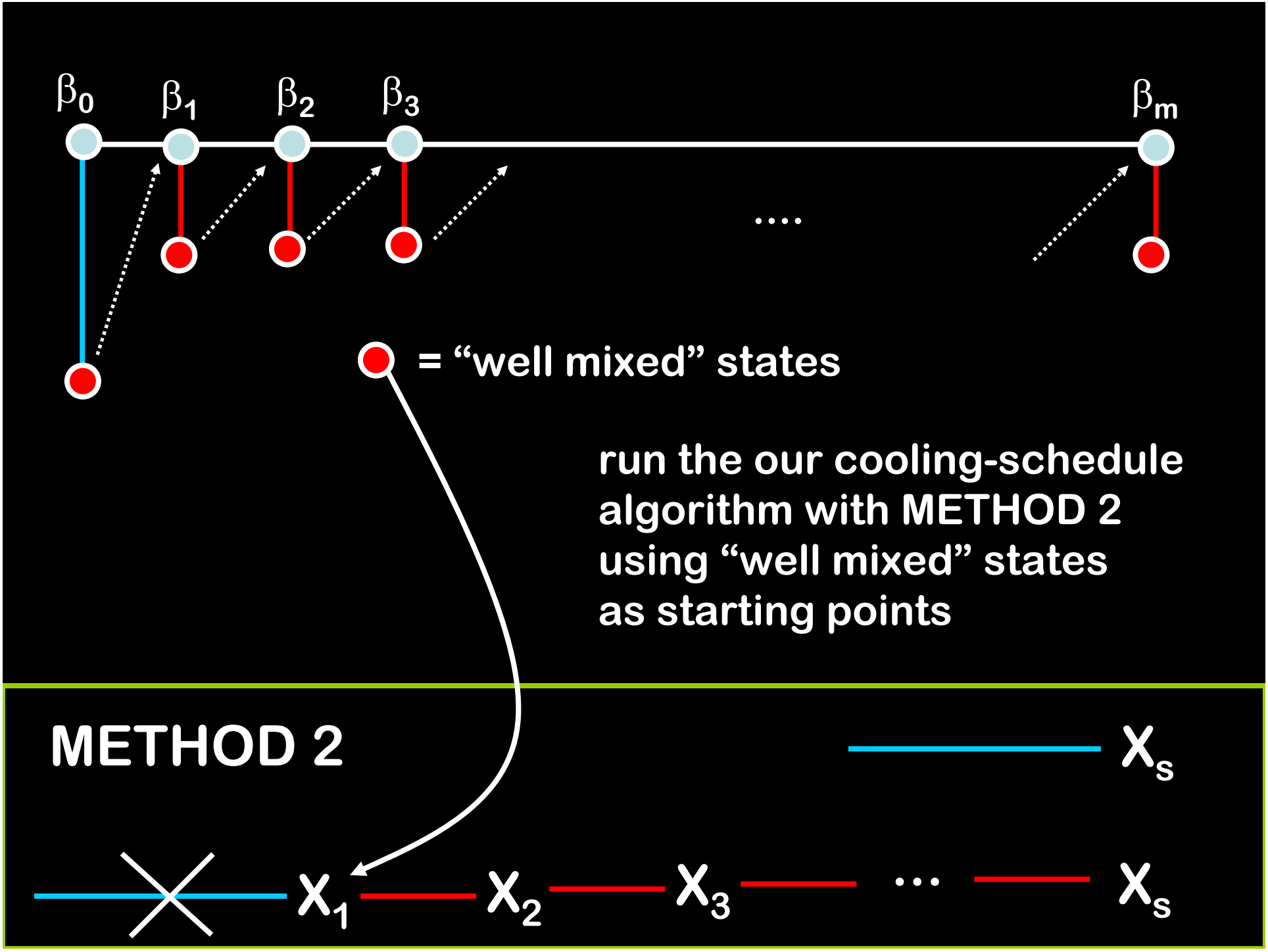
$O(n \ln A)$

$O((\ln n)(\ln A))$

(Bezáková,Štefankovič, Vigoda,V.Vazirani'06)

$m = O((\ln n)(\ln A))$

= "well mixed" states

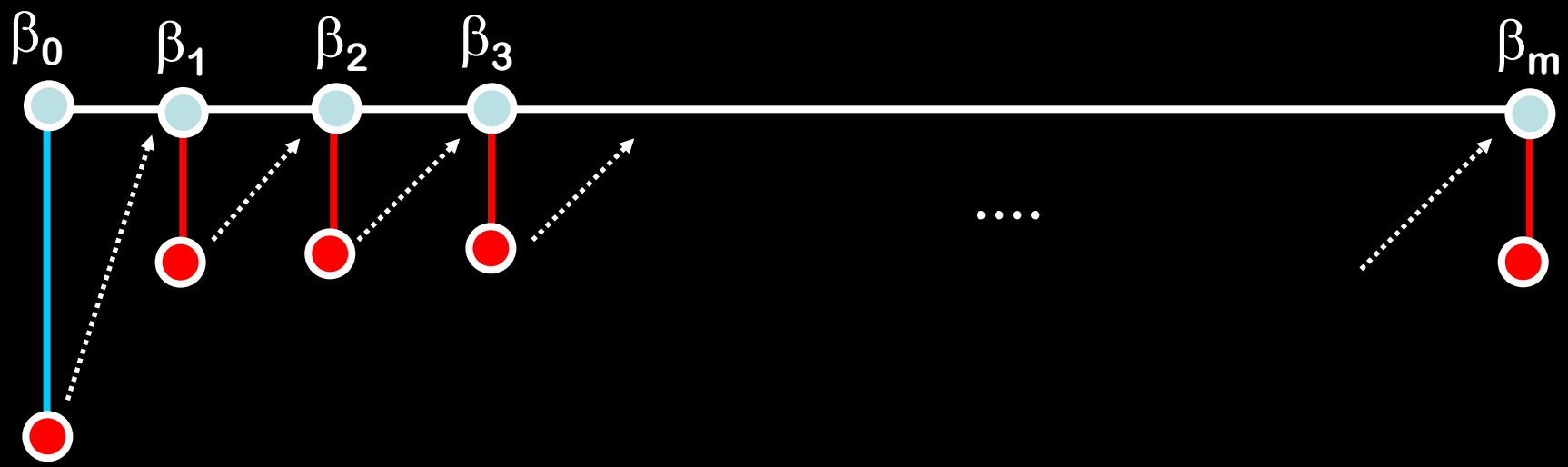

β0
β1
β2
β3
βm
....
= “well mixed” states
run the our cooling-schedule algorithm with METHOD 2 using “well mixed” states as starting points
METHOD 2
Xs
X1
X2
X3
...
Xs

Output of our algorithm:

$k=O^*((\ln A)^{1/2})$

small augmentation (so that we can use sample from current β as a warm start at next)

still $O^*((\ln A)^{1/2})$



Use analogue of Frieze-Dyer for independent samples from vector variables with slightly dependent coordinates.

if we have sampler oracles for μ_β then we can get adaptive schedule of length $t=O^*(\ (\ln A)^{1/2}\ )$

independent sets $O^*(n^2)$
(using Vigoda'01, Dyer-Greenhill'01)

matchings $O^*(n^2m)$
(using Jerrum, Sinclair'89)

spin systems:
Ising model $O^*(n^2)$ for $\beta<\beta_c$
(using Marinelli, Olivieri'95)
k-colorings $O^*(n^2)$ for $k>2\Delta$
(using Jerrum'95)