Approximate Counting via Correlation Decay

Pinyan Lu
Microsoft Research

Outline

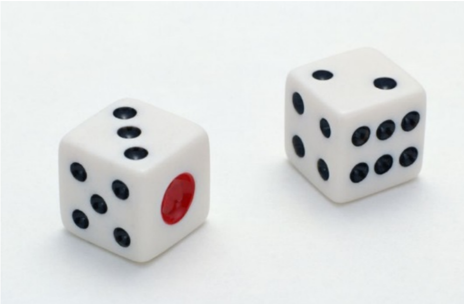
- Counting and probability distribution
- Two-state spin systems
- Multi-spin and Multi-part systems

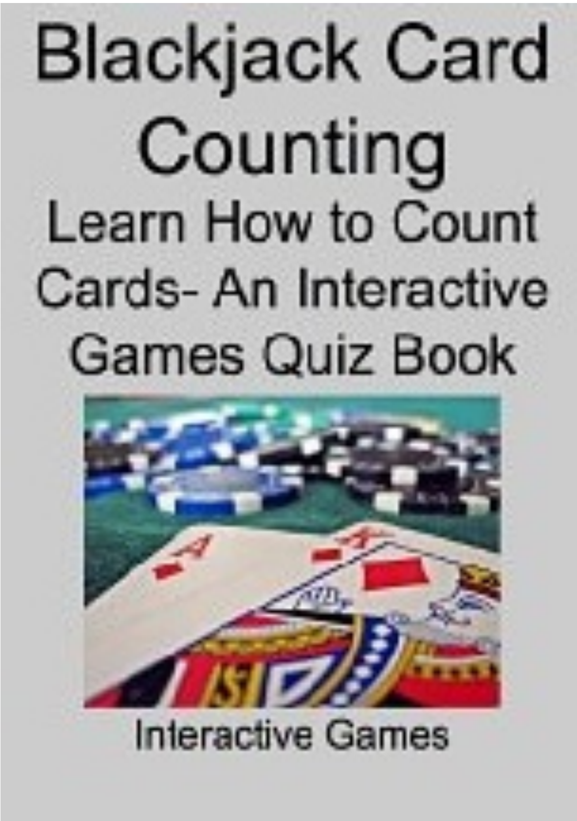
Counting Problems

- SAT: Is there a satisfying assignment for a given a CNF formula?
- Counting SAT: How many?
- Counting Colorings of a graph
- Counting Independent sets of a graph
- Counting perfect matchings of a bipartite graph (Permanent)
- ……

Counting Problems

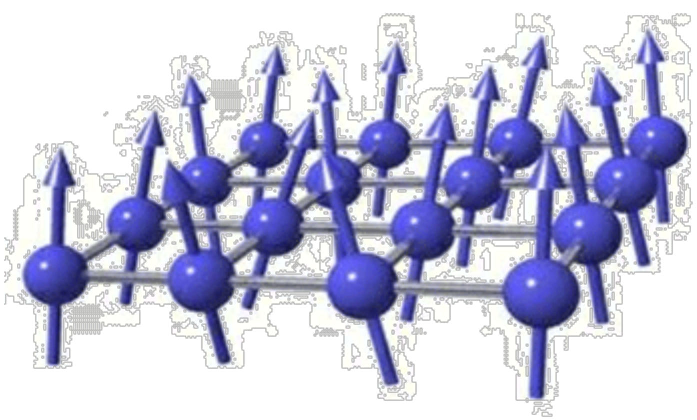
- Probability

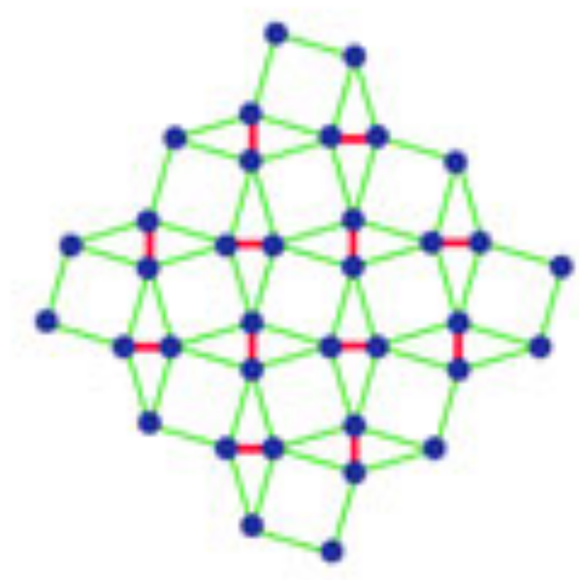




Counting Problems

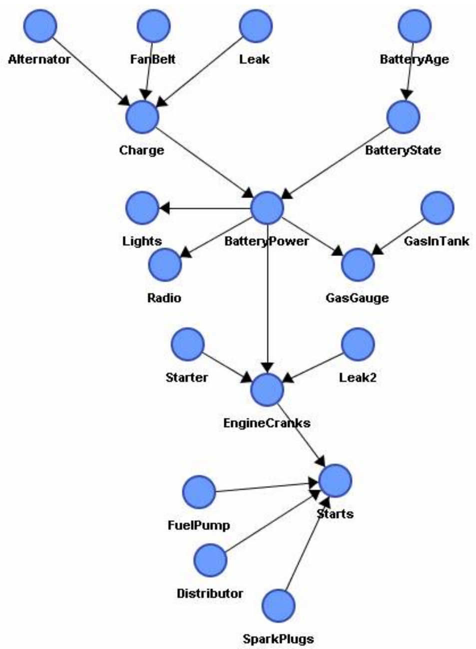
- Probability
- Partition function on statistical physics

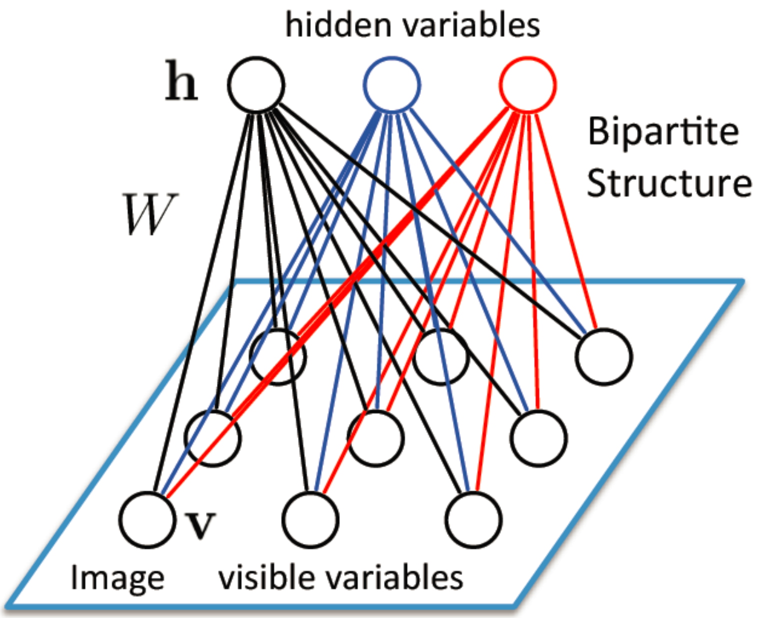




Counting Problems

- Probability
- Partition function on statistical physics
- Inference on Graphical Models





Counting Problems

- Probability
- Partition function on statistical physics
- Inference on Graphical Models
- Query on probabilistic database
- Optimization on stochastic model
- ……

Approximate Counting

- Let $\epsilon > 0$ be an approximation parameter and Z be the correct counting number of the instance, the algorithm returns a number $\hat{Z}$ such that $(1-\epsilon)Z \leq \hat{Z} \leq (1+\epsilon)Z$, in time ploy($n$,1/$\epsilon$).
- Fully polynomial-time approximation scheme (FPTAS).
- Fully polynomial-time randomized approximation scheme (FPRAS) is its randomized version.

Counting vs Distribution

- IS(G): the set of independent sets of graph G
- X is chosen from IS(G) uniformly at random
- $P_G(v)$: the probability that v is not in X
- $\Pr(X=\emptyset)=1/|IS(G)|$
- $\Pr(X=\emptyset)=P_{G_1}(v_1)P_{G_2}(v_2)\ldots P_{G_n}(v_n)$, where $G_1=G$, $G_{i+1}=G_i-v_i$

Counting vs Distribution

- $1/|IS(G)| = P_{G_1}(v_1) P_{G_2}(v_2) \ldots P_{G_n}(v_n)$
- If we can compute (estimate) $P_G(v)$, we can (approximately) compute |IS(G)|.

- FPRAS: Estimate $P_G(v)$ by sampling
- FPTAS: Approximately compute $P_G(v)$ directly and deterministically

Outline

- Counting and probability distribution
- Two-state spin systems
- Multi-spin and Multi-part systems

Spin Systems

- System $G=(V,E)$ and spin states $[q]$
- Configuration $\sigma:V\rightarrow[q]$
- Edge function $A:[q]\times[q]\rightarrow R\uparrow+$
- Vertex function b: [q]-> $R\uparrow+$
- Weight of a configuration

$w(\sigma)=(\prod(u,v)\in E\uparrow ▒ A(\sigma(u),\sigma(v))$
$(\prod v\in V\uparrow ▒ b(\sigma(v))\ )$

- Partition function: $Z\downarrow A(G)=\sum\sigma\uparrow ▒ w(\sigma)$

Gibbs Measure

- $\rho(\sigma)=w(\sigma)/Z_{A}(G)$ is a distribution over all configurations.
- We can define the marginal distribution of spins on a vertex p_{v}.
- We can also fix the configuration of a subset Λ of the vertices as σ_{Λ}, and define the conditional distribution of other vertex as $p_{v}^{\sigma_{\Lambda}}$.

Spin Systems

- A model in Statistical Physics and Applied Probability
- A framework of many combinatorial counting problems
- Have applications in AI, coding theory and so on.

Constraint Satisfaction Problems

- Graph coloring

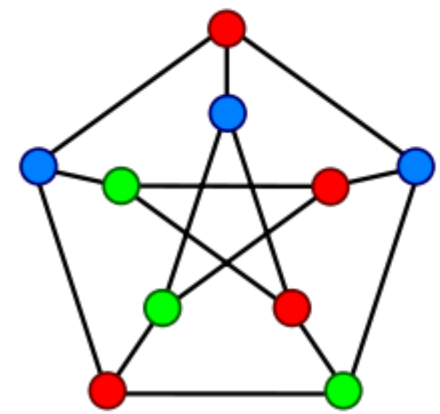
$$A = \begin{bmatrix} 0 & 1 & 1 \\ 1 & 0 & 1 \\ 1 & 1 & 0 \end{bmatrix}$$

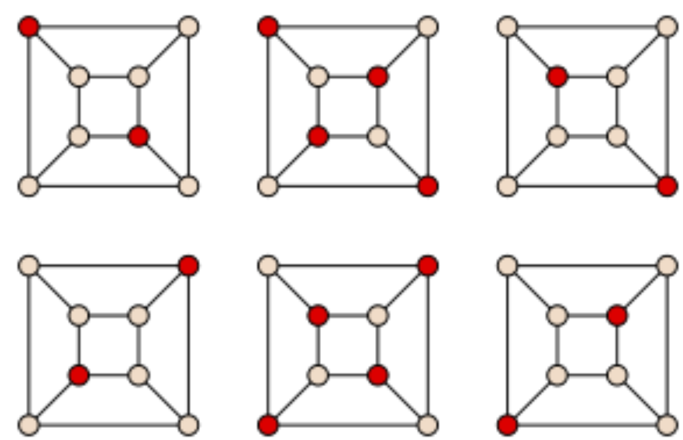
- Independent set

$$A = \begin{bmatrix} 1 & 1 \\ 1 & 0 \end{bmatrix}$$

Partition function:

Counting the number of solutions





Some Tasks

- Computing (conditional) marginal
- Sampling (wrt. the distribution)
- Computing partition function
- They are all related

Belief Propagation

- Also called message passing algorithms

- Widely used in statistical physics, machine learning and other applications

- Behave well in certain applications but without much theoretical justification unless the graph is a tree or very special

Belief Propagation

- Start with some initial guess
- Recursively refine the guess according to its local neighborhood
- Stop and output the current state after a fixed number of recursions
- Convergence and correctness?

Weak Correlation Decay

A spin system on a family of graphs is said to have exponential correlation decay if for any graph G=(V,E) in the family, any $v \in V, \Lambda \subset V$ and $\sigma_\Lambda, \tau_\Lambda \in \{0,1\}^\Lambda$,

$$|p_v^{\sigma_\Lambda} - p_v^{\tau_\Lambda}| \leq \exp(-\Omega(d(v,\Lambda))).$$

Correlation Decay

A spin system on a family of graphs is said to have exponential correlation decay if for any graph G=(V,E) in the family, any $v\in V, \Lambda\subset V$ and $\sigma_\Lambda, \tau_\Lambda \in\{0,1\}^\Lambda$,

$$|p_v^{\sigma_\Lambda} - p_v^{\tau_\Lambda}| \le \exp(-\Omega(d(v,S))),$$

where $S\subset\Lambda$ is the subset on which σ_Λ and τ_Λ differ.

Correlation Decay

- Significance of long distance interaction
- Effect of boundary condition
- Effect of the initial guess in Belief Propagation

- The big question: Characterize systems with correlation decay

Two-State Spin System

- After normalization: $A=[■\beta\&1@1\&\gamma]$, $b=[■\lambda 1]$
- Anti-ferromagnetic system: $\beta\gamma<1$

- Hardcore model : $A=[■1\&1@1\&0]$
- Ising model: $A=[■\beta\&1@1\&\beta]$

Uniqueness Condition

- $f(x)=\lambda(\beta x+1/x+\gamma)\uparrow d$.
- Let $x=f(x)$ be the fixed point of f.
- The system is called d-uniqueness if

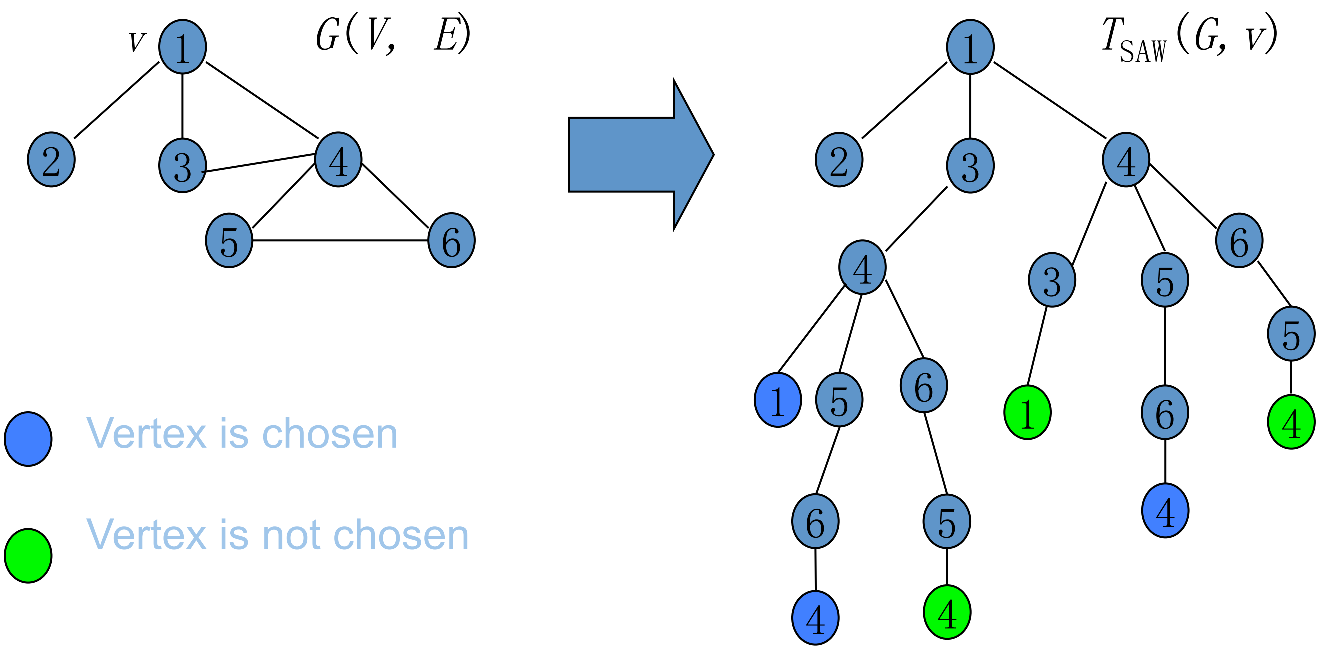
$|f\uparrow' (x)|<1$

- This can be numerically tested, but there is no closed form in general.

Hardcore Model [Weitz 2006]

- Strong correlation decay holds on all graphs with maximum degree at most Δ iff the uniqueness condition holds on infinite Δ -regular tree.
- Self Avoiding walk(SAW) tree: transform a general graph to a tree and the keep the marginal distribution for the root.
- Monotonicity: any tree with degree at most Δ decays at least as fast as the complete Δ -regular tree.

Self Avoiding Walk Tree



Self Avoiding Walk Tree

- It is enough to prove correlation decay on trees.
- The entire SAW tree is of exponential size comparing to the original graph.
- It only works for two-state spin systems.

Ising Model

[Sinclair, Srivastava, and Thurley 2011]

- Strong correlation decay holds on all graphs with maximum degree at most Δ iff the uniqueness condition holds on infinite Δ -regular tree.
- Have the same monotonicity property as hardcore model.

Non-Monotonicity

- The monotonicity does not hold for general two-state spin systems

- We need to prove correlation decay for general trees (with degrees up to Δ) rather than regular trees

- The previous techniques cannot be used

Our Results [Li, L., Yin 2012,13]

- The system is of correlation decay on all the graphs with maximum degree Δ iff the system exhibits uniqueness on all the infinite regular trees up to degree Δ.
- In particular, if the system exhibits uniqueness on infinite regular trees of all degrees, then the system is of correlation decay on all graphs.

Our Results [Li, L., Yin 2012,13]

- We obtain a FPTAS as long as the system satisfies the uniqueness condition.

- There is a matching hardness result [Sly,Sun 2012]: It is NP-hard if the system does not satisfy the uniqueness condition.

From correlation decay to FPTAS

- Marginal distribution $p_v^{\sigma_\Lambda}$ -> partition function

  $$w(0)/z = p_{v_1} \; p_{v_2}^{\sigma(1)=0} \ldots p_{v_n}^{\sigma(i)=0, i=1,2,\ldots,n-1}$$

- Correlation decay-> estimate $p_v^{\sigma_\Lambda}$ by a local neighborhood: $O(\log n)$ depth of the SAW tree.

- How about unbounded degree?

Computational Efficient Correlation Decay

- M-based depth:
  - $L\downarrow M\ (root)=0$;
  - $L\downarrow M\ (u)=L\downarrow M\ (v)+\lceil \log\downarrow M\ (d+1)\ \rceil$, if u is one of the d children of v.
- Exponential correlation decay with respect to M-based depth.
- Computational efficient correlation decay supports FPTAS for general graph.

Proof Sketch for Correlation Decay

- Self avoiding walk tree and recursion relation on tree: $p_v = f(p_{v_1}, p_{v_2}, \ldots, p_{v_d})$

- Estimate the error for one recursive step:

$$\epsilon_v = \frac{\partial f}{\partial p_{v_1}} \epsilon_{v_1} + \frac{\partial f}{\partial p_{v_2}} \epsilon_{v_2} + \ldots + \frac{\partial f}{\partial p_{v_d}} \epsilon_{v_d}$$

$$|\epsilon_v| \leq \left(\left|\frac{\partial f}{\partial p_{v_1}}\right| + \left|\frac{\partial f}{\partial p_{v_2}}\right| + \ldots + \left|\frac{\partial f}{\partial p_{v_d}}\right| \right) \max(|\epsilon_{v_i}|)$$

- $\left(\left|\frac{\partial f}{\partial p_{v_1}}\right| + \left|\frac{\partial f}{\partial p_{v_2}}\right| + \ldots + \left|\frac{\partial f}{\partial p_{v_d}}\right| \right) < 1$

Potential Function

- This may not be correct stepwise. We use a potential function to amortize it.
- Let $\phi: R^{+} \to R^{+}$ be a bijective function. $q_v = \phi(p_v), q_{v_i} = \phi(p_{v_i})$.
- $q_v = \phi(f(\phi^{-1}(q_{v_1}), \phi^{-1}(q_{v_2}), \ldots, \phi^{-1}(q_{v_d})))$.
- Then we show that the error for q is stepwise decreased by a constant factor.
- The main difficulty is to find the potential function ϕ.

A mathematical problem

- $f: R^d \to R^d$
- Contraction: $\forall\ x, \delta \in R^d$, we have $|J_f(x)\delta| \leq c\,|\delta|$ for some constant $c < 1$
- For some bijective mapping $\phi: R^d \to R^d$, such that $\phi \cdot f \cdot \phi^{-1}$ is contracted
- The effect of ϕ can be viewed as a local Riemann metric

Outline

- Counting and probability distribution
- Two-state spin systems
- Multi-spin and Multi-part systems

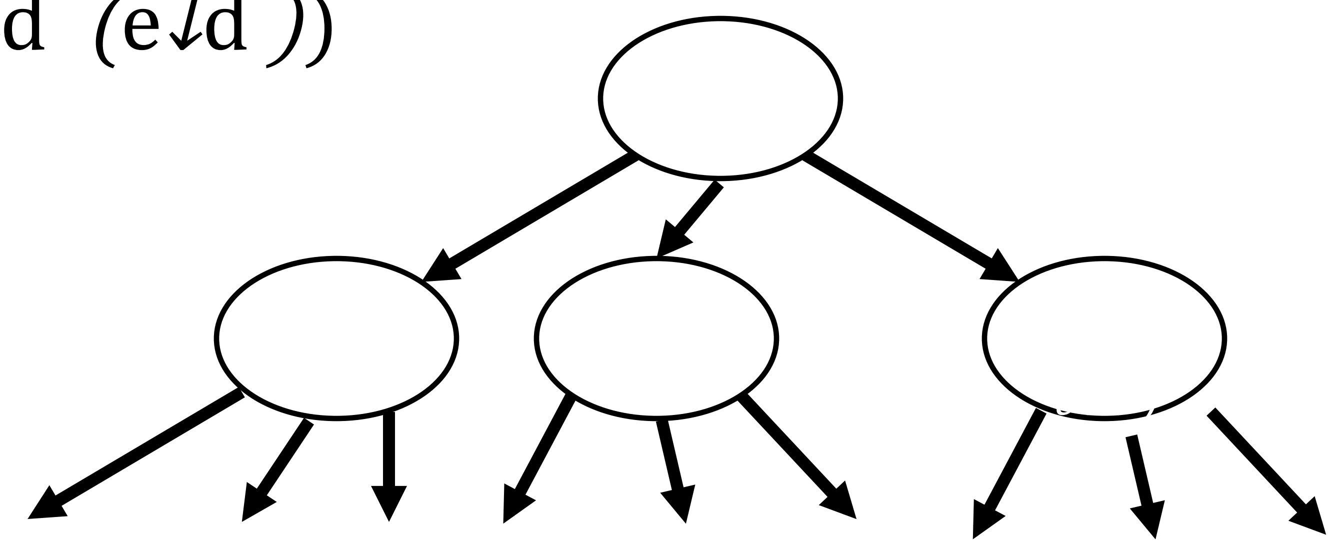
Multi-spin and Multi-part systems

- Multi-spin: the domain of the variables is larger than two
- Multi-part: each constraint involves more than two variables
- Obstacle: SAW tree does not work

Computation Tree

- Relate the probability $P_G(e)$ to these of its neighbors in smaller instances.

- $P_G(e)=f(P_{G_1}(e_1), P_{G_2}(e_2), \ldots, P_{G_d}(e_d))$



Correlation Decay

Truncate the computation tree at depth L, we can compute an estimation $P_G^L(e)$.

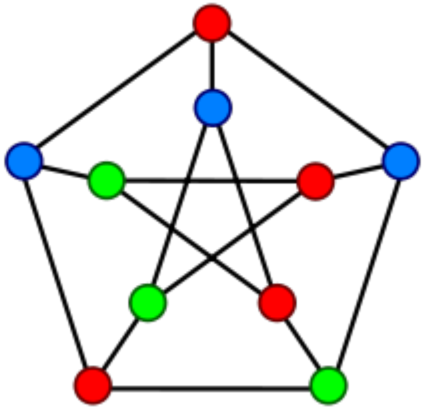
The system is called of exponential correlation decay if

$$|P_G^L(e) - P_G(e)| \leq \exp(-L).$$

We can estimate $P_G(e)$ by set $L = O(\log n + \log 1/\epsilon)$

Coloring

- It is NP-hard if $q < \Delta$ and there is always a solution if $q \geq \Delta$.
- FPTAS if $q > 2.8432\Delta + C$ [Gamarnik, Katz 07]
- FPTAS if $q > 2.581\Delta + 1$ [L., Yin 2013]



q: number of colors

Δ: maximum degree of the graph

Multi-spin System

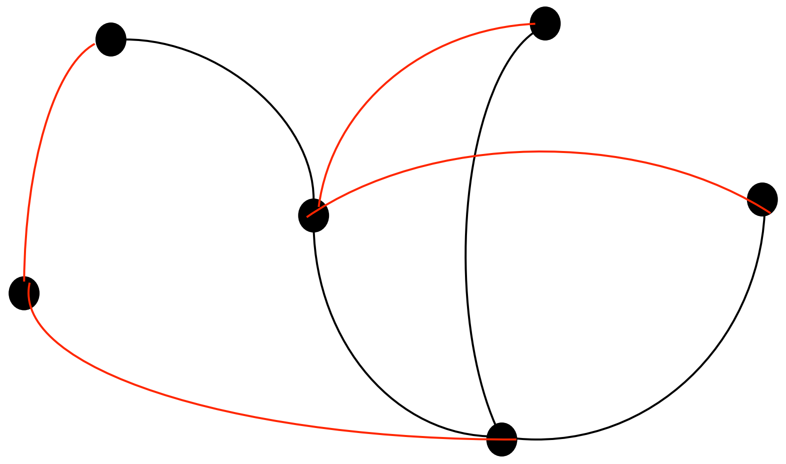
- Also called (Pairwise) Markov Random Field
- $c_A = \max A(x,y)/A(z,w)$
- We obtain an FPTAS when $3\Delta(c_A - 1) < 1$
- Potts model: inverse temperature $|\beta| = O(1/\Delta)$

Previously

- $(c_A^{\Delta} - c_A^{-\Delta})\Delta q^{\Delta} < 1$
- $|\beta| = O(1/\Delta q^{\Delta})$

Counting Edge Covers

- A set of edges such that every vertex has at least one adjacent edge in it



Counting Edge Covers

- A set of edges such that every vertex has at least one adjacent edge in it
- FPRAS for 3-regular graphs based on Markov Chain Monte Carlo[Bezakova, Rummler 2009].
- FPTAS for general graph. [Lin, Liu, L. 2014]

Counting Monotone CNF [Liu, L. 2015]

- A CNF formula is monotone if each variable appears positively.
- We give a FPTAS to count the number of solutions for a monotone CNF when each variable appears at most five times.
- It is NP-hard to approximately count if we allow a variable to appear six times.

Taking home messages

- Counting is in the heart of many computational problems, especially these related to probabilistic distribution.

- Correlation decay offers a new promising approach to design approximate counting algorithms.