Approximation via Correlation Decay when Strong Spatial Mixing Fails

Heng Guo

Queen Mary, University of London

Joint work with Ivona Bezáková, Andreas Galanis,
Leslie Ann Goldberg, and Daniel Štefankovič

Rome, Italy

Jul 13 2016

Independent sets

Graph $G = (V, E)$.

An independent set is a subset of V such that no two are adjacent.

$$\mathcal{I}(G) = \text{ the collection of independent sets in } G.$$

We are interested in approximating the size of $\mathcal{I}(G)$.

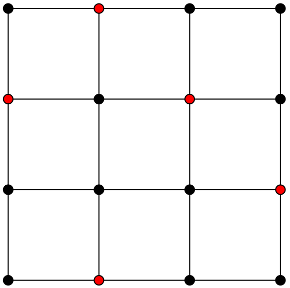
Independent sets

Graph $G = (V, E)$.

An independent set is a subset of V such that no two are adjacent.

$$\mathcal{I}(G) = \text{ the collection of independent sets in } G.$$

We are interested in approximating the size of $\mathcal{I}(G)$.



Independent

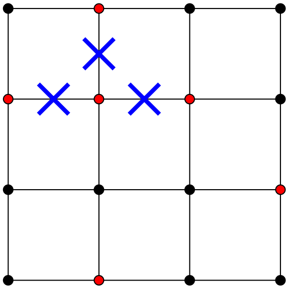
Independent sets

Graph $G = (V, E)$.

An independent set is a subset of V such that no two are adjacent.

$$\mathcal{I}(G) = \text{ the collection of independent sets in } G.$$

We are interested in approximating the size of $\mathcal{I}(G)$.



Not independent

The hardcore model

Consider the following distribution:

$$\pi(I) \propto \lambda^{|I|}$$

for $I \in \mathcal{I}(G)$ and some $\lambda > 0$.

This is also called the hardcore model with activity λ.

Partition function $Z = \sum_{I \in \mathcal{I}(G)} \lambda^{|I|}$.

In particular, if $\lambda = 1$, $Z = |\mathcal{I}(G)|$.

Computational transition

Approximate counting weighted independent sets

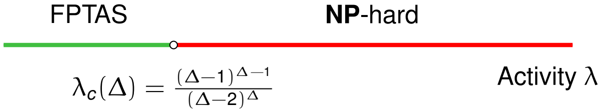
(or approximate Z for the hardcore model)

Computational transition

Approximate counting weighted independent sets

(or approximate Z for the hardcore model)

For G with a degree bound Δ:

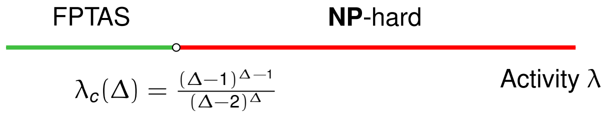


Computational transition

Approximate counting weighted independent sets

(or approximate Z for the hardcore model)

For G with a degree bound Δ:



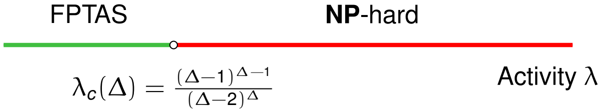
- Algorithm: [Weitz 06]

Computational transition

Approximate counting weighted independent sets

(or approximate Z for the hardcore model)

For G with a degree bound Δ:



- Algorithm: [Weitz 06]
- Hardness: [Sly 10] [Sly Sun 14] [Galanis, Štefankovič, Vigoda 16]

Counting independent sets

Specialize to approximate counting independent sets (fix $\lambda = 1$):

For G with a degree bound Δ:

FPTAS	**NP**-hard
$\Delta \leqslant 5$	$\Delta \geqslant 6$

($\Delta = 5$ is the largest integer so that $\lambda_c(\Delta) > 1$.)

- Algorithm: [Weitz 06]
- Hardness: [Sly 10]

Independent sets in hypergraphs

Hypergraph $H = (V, F)$, where a hyperedge $e \in F$ is a subset of V.

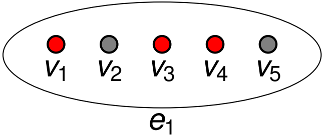
Independent set I: $I \subseteq V$ and $\forall e \in F$, $e \not\subset I$. (NOT-ALL-IN)

Independent sets in hypergraphs

Hypergraph $H = (V, F)$, where a hyperedge $e \in F$ is a subset of V.

Independent set I: $I \subseteq V$ and $\forall e \in F$, $e \not\subset I$. (NOT-ALL-IN)

Examples:



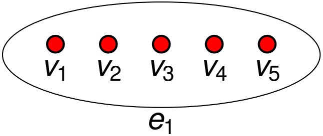
Independent

Independent sets in hypergraphs

Hypergraph $H = (V, F)$, where a hyperedge $e \in F$ is a subset of V.

Independent set I: $I \subseteq V$ and $\forall e \in F$, $e \not\subset I$. (NOT-ALL-IN)

Examples:



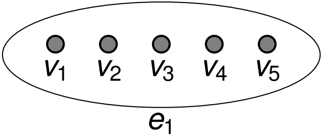
Not independent

Independent sets in hypergraphs

Hypergraph $H = (V, F)$, where a hyperedge $e \in F$ is a subset of V.

Independent set I: $I \subseteq V$ and $\forall e \in F$, $e \not\subset I$. (NOT-ALL-IN)

Examples:



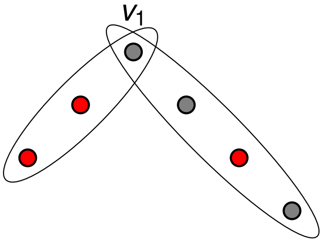
$2^5 - 1$ many independent sets

Independent sets in hypergraphs

Hypergraph $H = (V, F)$, where a hyperedge $e \in F$ is a subset of V.

Independent set I: $I \subseteq V$ and $\forall e \in F$, $e \not\subseteq I$. (NOT-ALL-IN)

Examples:



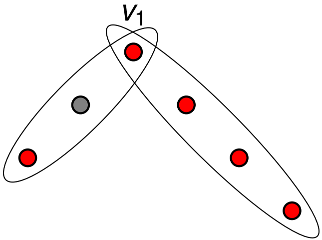
Independent

Independent sets in hypergraphs

Hypergraph $H = (V, F)$, where a hyperedge $e \in F$ is a subset of V.

Independent set I: $I \subseteq V$ and $\forall e \in F$, $e \not\subset I$. (NOT-ALL-IN)

Examples:



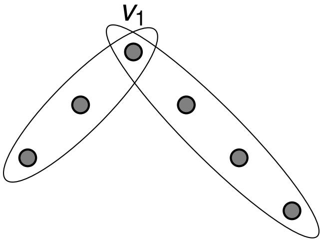
Not independent

Independent sets in hypergraphs

Hypergraph $H = (V, F)$, where a hyperedge $e \in F$ is a subset of V.

Independent set I: $I \subseteq V$ and $\forall e \in F$, $e \not\subset I$. (NOT-ALL-IN)

Examples:



$2^2 \cdot 2^3 + (2^2 - 1)(2^3 - 1)$

many independent sets

Monotone CNF

Independent Sets in Hypergraphs $\Leftrightarrow$

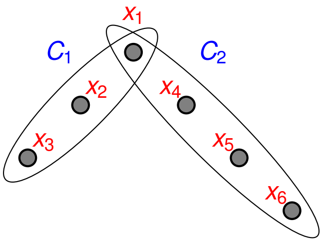
Satisfying assignments of monotone CNF formulas.

Monotone CNF

Independent Sets in Hypergraphs ⇔

Satisfying assignments of monotone CNF formulas.

Vertices are variables. Hyperedges are clauses.



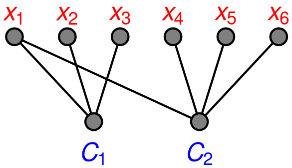
$$\underbrace{(x_1 \vee x_2 \vee x_3)}_{C_1} \wedge \underbrace{(x_1 \vee x_4 \vee x_5 \vee x_6)}_{C_2}$$

Monotone CNF

Independent Sets in Hypergraphs $\Leftrightarrow$

Satisfying assignments of monotone CNF formulas.

Vertices are variables. Hyperedges are clauses.



$$\underbrace{(x_1 \vee x_2 \vee x_3)}_{C_1} \wedge \underbrace{(x_1 \vee x_4 \vee x_5 \vee x_6)}_{C_2}$$

Bounded occurrences

Name #HyperIndSet(Δ, k).

Instance A hypergraph H with:

- maximum vertex degree $\leqslant \Delta$ (variable read-Δ);
- hyperedge cardinality $\geqslant k$ (clause arity).

Output The number Z of independent sets in H.

Previously ...

Based on Markov chain Monte Carlo:

- There is a FPRAS for #HYPERINDSET(Δ, k) if $k \geqslant \Delta + 2$.

  [Dyer Greenhill 00], [Borderwich, Dyer, Karpinski 08, 06].

Previously . . .

Based on Markov chain Monte Carlo:

- There is a FPRAS for #HyperIndSet(Δ, k) if $k \geqslant \Delta + 2$.
  [Dyer Greenhill 00], [Borderwich, Dyer, Karpinski 08, 06].

Based on correlation decay:

- There is a FPTAS for #HyperIndSet(Δ, k) if $\Delta \leqslant 5$ for any integer $k \geqslant 2$.
  [Liu Lu 15]

Previously ...

Based on Markov chain Monte Carlo:

- There is a FPRAS for #HyperIndSet(Δ, k) if $k \geqslant \Delta + 2$.
  [Dyer Greenhill 00], [Borderwich, Dyer, Karpinski 08, 06].

Based on correlation decay:

- There is a FPTAS for #HyperIndSet(Δ, k) if $\Delta \leqslant 5$ for any integer $k \geqslant 2$.
  [Liu Lu 15]

#HyperIndSet$(\Delta, 2)$ is at least as hard as counting independent sets.

Hence FPTAS for $k = 2$, $\Delta \leqslant 5$ is optimal. ($\Delta \geqslant 6$ is **NP**-hard [Sly 10].)

Our results

Theorem

There is a FPTAS for $\#\textsc{HyperIndSet}(\Delta, k)$ *if*

1. $\Delta = 6$ *and* $k \geqslant 3$;
2. *For* $\Delta \geqslant 200$, $k \geqslant 1.66\Delta$.

Our results

Theorem

There is a FPTAS for #HyperIndSet(Δ, k) *if*

1. $\Delta = 6$ *and* $k \geqslant 3$;
2. *For* $\Delta \geqslant 200$, ~~$k \geqslant 1.66\Delta$~~ $k \geqslant \Delta$.

Our results

Theorem

There is a FPTAS for #HYPERINDSET(Δ, k) if

1. $\Delta = 6$ *and* $k \geqslant 3$*;*
2. *For* $\Delta \geqslant 200$, ~~$k \geqslant 1.66\Delta$~~ $k \geqslant \Delta$.

- For $\Delta = 6$, $k = 3$ is optimal as #HYPERINDSET$(6, 2)$ is **NP**-hard [Sly 10].
- $k \geqslant \Delta$ only slightly improves $k \geqslant \Delta + 2$ [Borderwich, Dyer, Karpinski 06], but this improvement is essential for our application of counting dominating sets in regular graphs.

Hardness

Theorem

For any integer $\Delta \geqslant 5 \cdot 2^{k/2}$, it is **NP***-hard to approximate* #HYPERINDSET(Δ, k)*, even within an exponential factor.*

FPTAS: $\Delta \leqslant k$

NP-hard: $\Delta \geqslant 5 \cdot 2^{k/2}$

Dominating sets

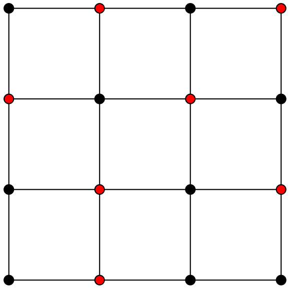
Graph $G = (V, E)$.

$D \subseteq V$ is dominating if every $v \in V$ either $\in D$ or is adjacent to some $v' \in D$.

Dominating sets

Graph $G = (V, E)$.

$D \subseteq V$ is dominating if every $v \in V$ either $\in D$ or is adjacent to some $v' \in D$.

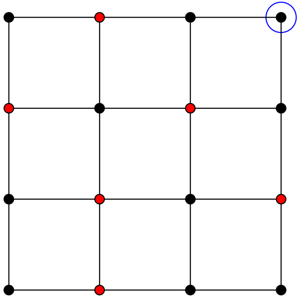


Dominating

Dominating sets

Graph $G = (V, E)$.

$D \subseteq V$ is dominating if every $v \in V$ either $\in D$ or is adjacent to some $v' \in D$.



Not dominating

Counting dominating sets

Name #RegDomSet(Δ).

Instance A Δ-regular graph G.

Output The number of dominating sets in G.

Dominating sets $\leqslant_T$ Independent sets in hypergraphs

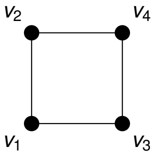
Express dominating sets as a CSP problem:

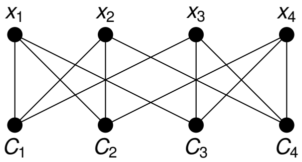
Each vertex is a variable and a constraint (NOT-ALL-OUT).

Dominating sets $\leqslant_T$ Independent sets in hypergraphs

Express dominating sets as a CSP problem:

Each vertex is a variable and a constraint (NOT-ALL-OUT).

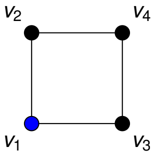


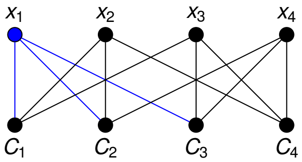


Dominating sets $\leqslant_T$ Independent sets in hypergraphs

Express dominating sets as a CSP problem:

Each vertex is a variable and a constraint (NOT-ALL-OUT).

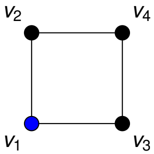


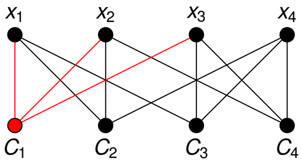


Dominating sets $\leqslant_T$ Independent sets in hypergraphs

Express dominating sets as a CSP problem:

Each vertex is a variable and a constraint (NOT-ALL-OUT).

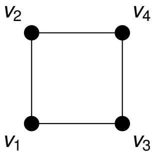


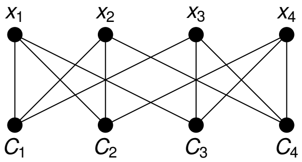


Dominating sets $\leqslant_{\mathrm{T}}$ Independent sets in hypergraphs

Express dominating sets as a CSP problem:

Each vertex is a variable and a constraint (NOT-ALL-OUT).





$$\#\text{REGDOMSET}(\Delta) \leqslant_T \#\text{HYPERINDSET}(\Delta + 1, \Delta + 1)$$

Counting dominating sets

$$\#\text{RegDomSet}(\Delta) \leqslant_T \#\text{HyperIndSet}(\Delta + 1, \Delta + 1)$$

Counting dominating sets

$$\#\textsc{RegDomSet}(\Delta) \leqslant_T \#\textsc{HyperIndSet}(\Delta+1, \Delta+1)$$

Theorem

There is a FPTAS for $\#\textsc{HyperIndSet}(\Delta, k)$ *if*

(1) $\Delta = 6$ and $k \geqslant 3$; (2) For $\Delta \geqslant 200$, $k \geqslant \Delta$.

Counting dominating sets

$$\#\textsc{RegDomSet}(\Delta) \leqslant_T \#\textsc{HyperIndSet}(\Delta+1, \Delta+1)$$

Theorem

There is a FPTAS for $\#\textsc{HyperIndSet}(\Delta, k)$ *if*

(1) $\Delta = 6$ and $k \geqslant 3$; (2) For $\Delta \geqslant 200$, $k \geqslant \Delta$.

Corollary

There is a FPTAS for $\#\textsc{RegDomSet}(\Delta)$ *if* $\Delta \leqslant 5$ *or* $\Delta \geqslant 199$*.*

Counting dominating sets

$$\#\text{RegDomSet}(\Delta) \leqslant_T \#\text{HyperIndSet}(\Delta+1, \Delta+1)$$

Theorem

There is a FPTAS for $\#\text{HyperIndSet}(\Delta, k)$ *if*

(1) $\Delta = 6$ and $k \geqslant 3$; (2) For $\Delta \geqslant 200$, $k \geqslant \Delta$.

Corollary

There is a FPTAS for $\#\text{RegDomSet}(\Delta)$ *if* $\Delta \leqslant 5$ *or* $\Delta \geqslant 199$.

Theorem

Approximately counting dominating sets is **NP**-*hard in graphs with bounded degree* $\Delta \geqslant 18$, *even within an exponential factor.*

Note the difference between being regular and bounded degree!

The Algorithm

A recursion for counting independent sets [Weitz 06]

$$P_G(v) = \frac{|\{I \in \mathcal{I} \mid v \notin I\}|}{|\mathcal{I}|} = \frac{Z(G-v)}{Z(G)}$$

$$= \frac{Z(G-v)}{Z(G-v) + Z(G-v-N(v))}$$

$$= \frac{1}{1 + \frac{Z(G-v-N(v))}{Z(G-v)}}$$

A recursion for counting independent sets [Weitz 06]

$$
\begin{aligned}
P_G(v) &= \frac{|\{I \in \mathcal{I} \mid v \notin I\}|}{|\mathcal{I}|} = \frac{Z(G-v)}{Z(G)} \\
&= \frac{Z(G-v)}{Z(G-v) + Z(G-v-N(v))} \\
&= \frac{1}{1 + \frac{Z(G-v-N(v))}{Z(G-v)}}
\end{aligned}
$$

Suppose $N(v) = \{v_1, \dots, v_d\}$.

$$
\frac{Z(G-v-N(v))}{Z(G-v)} = \frac{Z(G-v-v_1)}{Z(G-v)} \cdot \frac{Z(G-v-v_1-v_2)}{Z(G-v-v_1)} \cdots \frac{Z(G-v-N(v))}{Z(G-v-(N(v)-v_d))}
$$

A recursion for counting independent sets [Weitz 06]

$$\boxed{P_G(v) = \frac{|\{I \in \mathcal{I} \mid v \in I\}|}{|\mathcal{I}|} = \frac{Z(G-v)}{Z(G)}}$$

$$= \frac{Z(G-v)}{Z(G-v) + Z(G-v-N(v))}$$

$$= \frac{1}{1 + \frac{Z(G-v-N(v))}{Z(G-v)}}$$

Suppose $N(v) = \{v_1, \ldots, v_d\}$.

$$\frac{Z(G-v-N(v))}{Z(G-v)} = \frac{Z(G-v-v_1)}{Z(G-v)} \cdot \frac{Z(G-v-v_1-v_2)}{Z(G-v-v_1)} \cdot \cdots \cdot \frac{Z(G-v-N(v))}{Z(G-v-(N(v)-v_d))}$$

$$= P_{G_1}(v_1) \cdot P_{G_2}(v_2) \cdot \cdots \cdot P_{G_d}(v_d)$$

Here $G_i = G - v - v_1 - \cdots - v_{i-1}$.

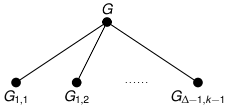
Computation Tree

The algorithm for $\#\textsc{HyperIndSet}(\Delta, k)$ is a similar recursion [Liu Lu 15], except that each step has possibly $(\Delta - 1)(k - 1)$ many branches.

Computation Tree

The algorithm for #HyperIndSet(Δ, k) is a similar recursion [Liu Lu 15], except that each step has possibly $(\Delta - 1)(k - 1)$ many branches.

- The recursion forms a computation tree.
- We stop the recursion after $O(\log n)$ many steps.
- The layer of depth ℓ corresponds to vertices that have distance ℓ from v.
- The main task is to bound the error.



Strong spatial mixing

SSM: Let σ_Λ and τ_Λ be two partial configurations on $\Lambda \subseteq V$.

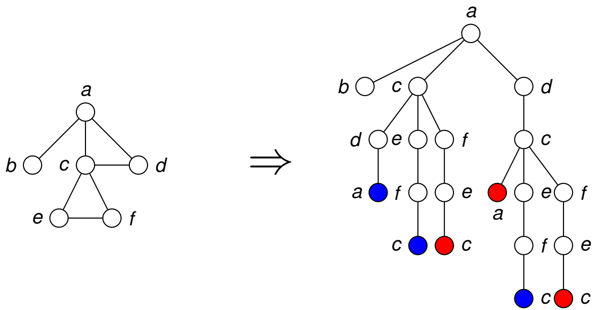
Let S be the set where σ_Λ and τ_Λ differ.

$$|p_v^{\sigma_\Lambda} - p_v^{\tau_\Lambda}| \leqslant \exp(-\Omega(\mathrm{dist}(v, S)))$$

Roughly speaking, the influence of the boundary decays exponentially,

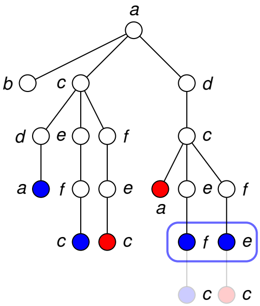
even with some vertices fixed within the radius.

An example: recursion

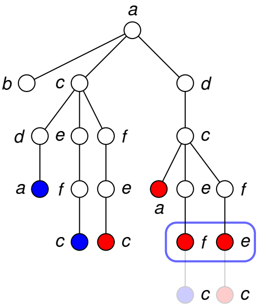


An example: strong spatial mixing

SSM:



V.S.



SSM for hypergraph independent sets

In the computation tree, hyperedge sizes will decrease.

Eventually, the size may go down to 2.

Hence SSM does not hold for independent sets in hypergraphs when $\Delta \geqslant 6$.

(SSM does not hold for independent sets in graphs if $\Delta \geqslant 6$.)

This is why [Liu, Lu 15] can only do $\Delta \leqslant 5$.

Beyond strong spatial mixing

Our contribution is to provide a way to analyze correlation decay beyond the strong spatial mixing bound.

- Larger hyperedges have better decay.
- Keep track of the total “deficits” of sub-instances.
- Amortized analysis — SSM is worst case.

Main difficulty — bound the decay rate

In the technical level, the main difficulty is to bound the decay rate function — a optimisation problem with possibly $(\Delta - 1)(k - 1)$ many variables.

Main difficulty — bound the decay rate

In the technical level, the main difficulty is to bound the decay rate function — a optimisation problem with possibly $(\Delta-1)(k-1)$ many variables.

Decay Rate

$$\kappa^{d,\mathbf{k}}(\mathbf{r}) := \frac{1}{\psi - F(\mathbf{r})^{\chi}} \sum_{i=1}^{d} \alpha^{-l_{k_i-1}} \frac{\prod_{j=1}^{k_i-1} \frac{r_{i,j}}{1+r_{i,j}}}{1-\prod_{j=1}^{k_i-1} \frac{r_{i,j}}{1+r_{i,j}}} \sum_{j=1}^{k_i-1} \delta^{c_{i,j}} \frac{\psi - r_{i,j}^{\chi}}{1+r_{i,j}},$$

where

$$F(\mathbf{r}) = \prod_{i=1}^{d} \left(1 - \prod_{j=1}^{k_i-1} \frac{r_{i,j}}{1+r_{i,j}}\right)$$

$$c_{i,j} = b_2(k-2) + s_{\min(i,d-b_2)} - \max(0, b'_k - i) - (j-1)(\Delta-1)\mathbf{1}_{i\leqslant d-b_2}.$$

Open questions

- The exact threshold for #HyperIndSet(Δ, k)?
- Close the gap for #RegDomSet(Δ).
- Other instances where SSM fails to capture the complexity?

Open questions

- The exact threshold for #HyperIndSet(Δ, k)?
- Close the gap for #RegDomSet(Δ).
- Other instances where SSM fails to capture the complexity?

Thank You!

Full version: `arxiv.org/abs/1510.09193`