Local Markov Chains, Path Coupling and Belief Propagation (BP)

Eric Vigoda

Georgia Tech

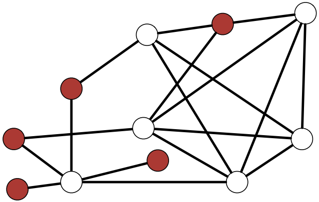
joint work with:

Charis Efthymiou (Warwick)
Tom Hayes (New Mexico)
Daniel Štefankovič (Rochester)
Yitong Yin (Nanjing)

WOLA, July '19

Independent Set

Undirected graph $G = (V, E)$:



Independent set: subset of vertices with no adjacent pairs.

Let $\Omega =$ all independent sets (of all sizes).

Our Goal:

1. *Counting problem:* Estimate $|\Omega|$.
2. *Sampling problem:* Sample uniformly at random from Ω.

Glauber dynamics = Gibbs Sampler

Given $G = (V, E)$, Markov chain (X_t) on Ω = all independent sets.

Transitions $X_t \to X_{t+1}$:

Glauber dynamics = Gibbs Sampler

Given $G = (V, E)$, Markov chain (X_t) on Ω = all independent sets.

Transitions $X_t \rightarrow X_{t+1}$:

1. Choose v uniformly at random from V.

Glauber dynamics = Gibbs Sampler

Given $G = (V, E)$, Markov chain (X_t) on $\Omega =$ all independent sets.

Transitions $X_t \to X_{t+1}$:

1. Choose v uniformly at random from V.

$$X' = \begin{cases} X_t \cup \{v\} & \text{with probability } 1/2 \\ X_t \setminus \{v\} & \text{with probability } 1/2 \end{cases}$$

Glauber dynamics = Gibbs Sampler

Given $G = (V, E)$, Markov chain (X_t) on Ω = all independent sets.

Transitions $X_t \to X_{t+1}$:

1. Choose v uniformly at random from V.

$$X' = \begin{cases} X_t \cup \{v\} & \text{with probability } 1/2 \\ X_t \setminus \{v\} & \text{with probability } 1/2 \end{cases}$$

2. If $X' \in \Omega$, then $X_{t+1} = X'$, otherwise $X_{t+1} = X_t$

Stationary distribution is $\mu = \text{uniform}(\Omega)$.

Mixing Time: $T_{\rm mix} := \min\{t : \text{ for all } X_0,\ d_{tv}(X_t, \mu) \le 1/4\ \}$

Then $T_{\rm mix}(\epsilon) \le \log(1/\epsilon) T_{\rm mix}$.

Recall, $d_{\text{TV}}(\mu, \nu) = \frac{1}{2} \sum_{\sigma \in \Omega} |\mu(\sigma) - \nu(\sigma)|$.

Independent Sets

Given input graph $G = (V, E)$ with $n = |V|$ vertices,
let $\Omega =$ set of all independent sets in G.

Typically, $|\Omega|$ is HUGE = exponentially large in n.

Independent Sets

Given input graph $G = (V, E)$ with $n = |V|$ vertices,
let $\Omega =$ set of all independent sets in G.

Typically, $|\Omega|$ is HUGE $=$ exponentially large in n.

Goal: in time poly(n):

1. Counting: Compute $|\Omega|$,
2. Sampling: generate random element of Ω.

Independent Sets

Given input graph $G = (V, E)$ with $n = |V|$ vertices,
let $\Omega =$ set of all independent sets in G.

Typically, $|\Omega|$ is HUGE = exponentially large in n.

Goal: in time poly(n):

1. Counting: Compute $|\Omega|$,
2. Sampling: generate random element of Ω.

Exactly computing $|\Omega|$ is #P-complete,
even for maximum degree $\Delta = 3$. [Greenhill '00]

Approximate $|\Omega|$:

FPRAS for Z: Given G, $\epsilon, \delta > 0$, output EST where:

$$\mathbf{Pr}\left(\text{EST}(1-\epsilon) \le Z \le \text{EST}(1+\epsilon)\right) \ge 1-\delta,$$

in time poly$(|G|, 1/\epsilon, \log(1/\delta))$.

FPTAS for Z: FPRAS with $\delta = 0$.

Independent Sets

Given input graph $G = (V, E)$ with $n = |V|$ vertices,
let $\Omega =$ set of all independent sets in G.

Typically, $|\Omega|$ is HUGE = exponentially large in n.

Independent Sets

Given input graph $G = (V, E)$ with $n = |V|$ vertices,
let Ω = set of all independent sets in G.

Typically, $|\Omega|$ is HUGE = exponentially large in n.

Goal: in time poly(n):

Compute $|\Omega|$ or sample from uniform(Ω)?

Independent Sets

Given input graph $G = (V, E)$ with $n = |V|$ vertices,
let $\Omega =$ set of all independent sets in G.

Typically, $|\Omega|$ is HUGE = exponentially large in n.

Goal: in time poly(n):
Compute $|\Omega|$ or sample from uniform(Ω)?

General graphs:
NP-hard to approx. $|\Omega|$ within $2^{n^{1-\epsilon}}$ for any $\epsilon > 0$.

Independent Sets

Given input graph $G = (V, E)$ with $n = |V|$ vertices,
let $\Omega =$ set of all independent sets in G.

Typically, $|\Omega|$ is HUGE = exponentially large in n.

Goal: in time poly(n):
Compute $|\Omega|$ or sample from uniform(Ω)?

General graphs:
NP-hard to approx. $|\Omega|$ within $2^{n^{1-\epsilon}}$ for any $\epsilon > 0$.

Restricted graphs: Given graph G with maximum degree Δ:
For $\Delta \leq 5$, FPTAS for $|\Omega|$. [Weitz '06]

For $\Delta \geq 6$, $\exists \delta > 0$, no poly-time to approx $|\Omega|$ within $2^{n^{\delta}}$ unless $NP = RP$. [Sly '10]

Independent Sets

Given input graph $G = (V, E)$ with $n = |V|$ vertices,
let Ω = set of all independent sets in G.

Typically, $|\Omega|$ is HUGE = exponentially large in n.

Goal: in time poly(n):
Compute $|\Omega|$ or sample from uniform(Ω)?

General graphs:
NP-hard to approx. $|\Omega|$ within $2^{n^{1-\epsilon}}$ for any $\epsilon > 0$.

Restricted graphs: Given graph G with maximum degree Δ:
For $\Delta \leq 5$, FPTAS for $|\Omega|$. [Weitz '06]

For $\Delta \geq 6$, $\exists \delta > 0$, no poly-time to approx $|\Omega|$ within $2^{n^{\delta}}$ unless $NP = RP$. [Sly '10]

What happens between $\Delta = 5 \leftrightarrow 6$?
Statistical physics phase transition on infinite Δ-regular tree!

Hard-Core Gas Model

Graph $G = (V, E)$, *fugacity* $\lambda > 0$, for $\sigma \in \Omega$:

Gibbs distribution: $$\mu(\sigma) = \frac{\lambda^{|\sigma|}}{Z}$$

where

Partition function: $$Z = \sum_{\sigma} \lambda^{|\sigma|}$$

$\lambda = 1$, $Z = |\Omega| = \#$ of independent sets.

Hard-Core Gas Model

Graph $G = (V, E)$, *fugacity* $\lambda > 0$, for $\sigma \in \Omega$:

Gibbs distribution: $$\mu(\sigma) = \frac{\lambda^{|\sigma|}}{Z}$$

where

Partition function: $$Z = \sum_{\sigma} \lambda^{|\sigma|}$$

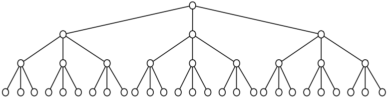
$\lambda = 1$, $Z = |\Omega| = \#$ of independent sets.

Inuition: Small λ easier: for $\lambda < 1$ prefer smaller sets.
Large λ harder: for $\lambda > 1$ prefer max independent sets.

Phase Transition on Trees

For Δ-regular tree of height ℓ:

$$\text{Let } p_\ell := \mathbf{Pr}\,(\text{root is occupied})$$

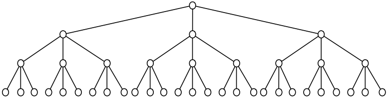


Extremal cases: even versus odd height.

Phase Transition on Trees

For Δ-regular tree of height ℓ:

$$\text{Let } p_\ell := \mathbf{Pr}\,(\text{root is occupied})$$



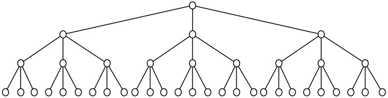
Extremal cases: even versus odd height.

$$\text{Does } \lim_{\ell\to\infty} p_{2\ell} = \lim_{\ell\to\infty} p_{2\ell+1} \text{ ?}$$

Phase Transition on Trees

For Δ-regular tree of height ℓ:

$$\text{Let } p_\ell := \mathbf{Pr}(\text{root is occupied})$$



Extremal cases: even versus odd height.

$$\text{Does } \lim_{\ell\to\infty} p_{2\ell} = \lim_{\ell\to\infty} p_{2\ell+1} \text{ ?}$$

$\lambda_c(\Delta) = \frac{(\Delta-1)^{\Delta-1}}{(\Delta-2)^{\Delta}} \approx \frac{e}{\Delta-2}$.

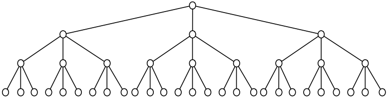
$\lambda \leq \lambda_c(\Delta)$: No boundary affects root.

$\lambda > \lambda_c(\Delta)$: Exist boundaries affect root.

Phase Transition on Trees

For Δ-regular tree of height ℓ:

$$\text{Let } p_\ell := \mathbf{Pr}\,(\text{root is occupied})$$



Extremal cases: even versus odd height.

$$\text{Does } \lim_{\ell\to\infty} p_{2\ell} = \lim_{\ell\to\infty} p_{2\ell+1} \text{ ?}$$

$\lambda_c(\Delta) = \frac{(\Delta-1)^{\Delta-1}}{(\Delta-2)^{\Delta}} \approx \frac{e}{\Delta-2}$.

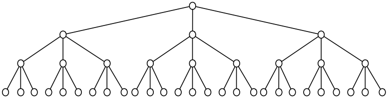
$\lambda \leq \lambda_c(\Delta)$: No boundary affects root. uniqueness

$\lambda > \lambda_c(\Delta)$: Exist boundaries affect root. non-uniqueness

Phase Transition on Trees

For Δ-regular tree of height ℓ:

$$\text{Let } p_\ell := \mathbf{Pr}\,(\text{root is occupied})$$



Extremal cases: even versus odd height.

$$\text{Does } \lim_{\ell\to\infty} p_{2\ell} = \lim_{\ell\to\infty} p_{2\ell+1} \text{ ?}$$

$\lambda_c(\Delta) = \frac{(\Delta-1)^{\Delta-1}}{(\Delta-2)^{\Delta}} \approx \frac{e}{\Delta-2}$.

$\lambda \leq \lambda_c(\Delta)$: No boundary affects root. uniqueness

$\lambda > \lambda_c(\Delta)$: Exist boundaries affect root. non-uniqueness

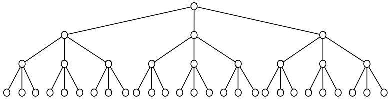
Example: $\Delta = 5$, $\lambda = 1$:

$$p_{even} = .245, \ p_{odd} = .245$$

Phase Transition on Trees

For Δ-regular tree of height ℓ:

$$\text{Let } p_\ell := \mathbf{Pr}\,(\text{root is occupied})$$



Extremal cases: even versus odd height.

$$\text{Does } \lim_{\ell\to\infty} p_{2\ell} = \lim_{\ell\to\infty} p_{2\ell+1} \text{ ?}$$

$\lambda_c(\Delta) = \frac{(\Delta-1)^{\Delta-1}}{(\Delta-2)^{\Delta}} \approx \frac{e}{\Delta-2}$.

$\lambda \le \lambda_c(\Delta)$: No boundary affects root. uniqueness

$\lambda > \lambda_c(\Delta)$: Exist boundaries affect root. non-uniqueness

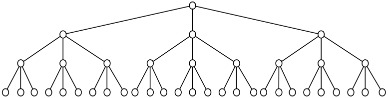
Example: $\Delta = 5$, $\lambda = 1.05$:

$$p_{even} = .250, \;\; p_{odd} = .250$$

Phase Transition on Trees

For Δ-regular tree of height ℓ:

$$\text{Let } p_\ell := \mathbf{Pr}\,(\text{root is occupied})$$



Extremal cases: even versus odd height.

$$\text{Does } \lim_{\ell\to\infty} p_{2\ell} = \lim_{\ell\to\infty} p_{2\ell+1} \text{ ?}$$

$\lambda_c(\Delta) = \frac{(\Delta-1)^{\Delta-1}}{(\Delta-2)^{\Delta}} \approx \frac{e}{\Delta-2}$.

$\lambda \le \lambda_c(\Delta)$: No boundary affects root. uniqueness

$\lambda > \lambda_c(\Delta)$: Exist boundaries affect root. non-uniqueness

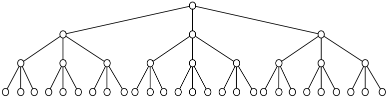
Example: $\Delta = 5$, $\lambda = 1.06$:

$$p_{even} = .283, \; p_{odd} = .219$$

Phase Transition on Trees

For Δ-regular tree of height ℓ:

Let $p_\ell := \mathbf{Pr}\,(\text{root is occupied})$



Extremal cases: even versus odd height.

Does $\lim_{\ell\to\infty} p_{2\ell} = \lim_{\ell\to\infty} p_{2\ell+1}$?

$\lambda_c(\Delta) = \frac{(\Delta-1)^{\Delta-1}}{(\Delta-2)^{\Delta}} \approx \frac{e}{\Delta-2}$.

$\lambda \le \lambda_c(\Delta)$: No boundary affects root. uniqueness

$\lambda > \lambda_c(\Delta)$: Exist boundaries affect root. non-uniqueness

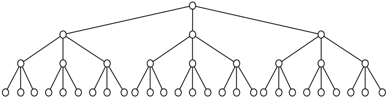
Tree/BP recursions: $p_{\ell+1} = \frac{\lambda(1-p_\ell)^{\Delta-1}}{1+\lambda(1-p_\ell)^{\Delta-1}}$

Key: Unique vs. Multiple fixed points of 2-level recursions.

Phase Transition on Trees

For Δ-regular tree of height ℓ:

$$\text{Let } p_\ell := \mathbf{Pr}\,(\text{root is occupied})$$



Extremal cases: even versus odd height.

$$\text{Does } \lim_{\ell\to\infty} p_{2\ell} = \lim_{\ell\to\infty} p_{2\ell+1} \text{ ?}$$

$\lambda_c(\Delta) = \frac{(\Delta-1)^{\Delta-1}}{(\Delta-2)^{\Delta}} \approx \frac{e}{\Delta-2}$.

$\lambda \le \lambda_c(\Delta)$: No boundary affects root. uniqueness

$\lambda > \lambda_c(\Delta)$: Exist boundaries affect root. non-uniqueness

For 2-dimensional integer lattice $\mathbb{Z}^2$:

Conjecture:	$\lambda_c(\mathbb{Z}^2) \approx 3.79$
Best bounds:	$2.53 < \lambda_c(\mathbb{Z}^2) < 5.36$

Approximating Partition Function

Tree threshold: $\lambda_c(\Delta) := \frac{(\Delta-1)^{\Delta-1}}{(\Delta-2)^{\Delta}} \sim \frac{e}{\Delta}$:

Approximating Partition Function

Tree threshold: $\lambda_c(\Delta) := \frac{(\Delta-1)^{\Delta-1}}{(\Delta-2)^{\Delta}} \sim \frac{e}{\Delta}$:

- All constant Δ, all $\lambda < \lambda_c(\Delta)$, FPTAS for Z. [Weitz '06]

Approximating Partition Function

Tree threshold: $\lambda_c(\Delta) := \frac{(\Delta-1)^{\Delta-1}}{(\Delta-2)^{\Delta}} \sim \frac{e}{\Delta}$:

- All constant Δ, all $\lambda < \lambda_c(\Delta)$, FPTAS for Z. [Weitz '06]

Approximating Partition Function

Tree threshold: $\lambda_c(\Delta) := \frac{(\Delta-1)^{\Delta-1}}{(\Delta-2)^{\Delta}} \sim \frac{e}{\Delta}$:

- All constant Δ, all $\lambda < \lambda_c(\Delta)$, FPTAS for Z. [Weitz '06]

- All $\Delta \geq 3$, all $\lambda > \lambda_c(\Delta)$: No poly-time to approx. Z for Δ-regular, triangle-free G, unless $NP = RP$

[Sly '10,Galanis,Stefankovic,V '13, Sly,Sun '13, GSV '15]

Approximating Partition Function

Tree threshold: $\lambda_c(\Delta) := \frac{(\Delta-1)^{\Delta-1}}{(\Delta-2)^{\Delta}} \sim \frac{e}{\Delta}$:

- All constant Δ, all $\lambda < \lambda_c(\Delta)$, FPTAS for Z. [Weitz '06]
- FPTAS using Barvinok's approach. [Patel,Regts '17, Peters,Regts '19]

- All $\Delta \geq 3$, all $\lambda > \lambda_c(\Delta)$: No poly-time to approx. Z for Δ-regular, triangle-free G, unless $NP = RP$

[Sly '10,Galanis,Stefankovic,V '13, Sly,Sun '13, GSV '15]

Approximating Partition Function

Tree threshold: $\lambda_c(\Delta) := \frac{(\Delta-1)^{\Delta-1}}{(\Delta-2)^{\Delta}} \sim \frac{e}{\Delta}$:

- All constant Δ, all $\lambda < \lambda_c(\Delta)$, FPTAS for Z. [Weitz '06]
- FPTAS using Barvinok's approach. [Patel,Regts '17, Peters,Regts '19]

- All $\Delta \geq 3$, all $\lambda > \lambda_c(\Delta)$: No poly-time to approx. Z for Δ-regular, triangle-free G, unless $NP = RP$

[Sly '10,Galanis,Stefankovic,V '13, Sly,Sun '13, GSV '15]

What happens at $\lambda_c(\Delta)$?

Approximating Partition Function

Tree threshold: $\lambda_c(\Delta) := \frac{(\Delta-1)^{\Delta-1}}{(\Delta-2)^{\Delta}} \sim \frac{e}{\Delta}$:

- All constant Δ, all $\lambda < \lambda_c(\Delta)$, FPTAS for Z. [Weitz '06]
- FPTAS using Barvinok's approach. [Patel,Regts '17, Peters,Regts '19]

- All $\Delta \geq 3$, all $\lambda > \lambda_c(\Delta)$: No poly-time to approx. Z for Δ-regular, triangle-free G, unless $NP = RP$

[Sly '10,Galanis,Stefankovic,V '13, Sly,Sun '13, GSV '15]

What happens at $\lambda_c(\Delta)$?

Statistical physics phase transition on infinite Δ-regular tree

Computational phase transition on general max deg. Δ graphs

Approximating Partition Function

Tree threshold: $\lambda_c(\Delta) := \frac{(\Delta-1)^{\Delta-1}}{(\Delta-2)^{\Delta}} \sim \frac{e}{\Delta}$:

- All constant Δ, all $\lambda < \lambda_c(\Delta)$, FPTAS for Z. [Weitz '06]
- FPTAS using Barvinok's approach. [Patel,Regts '17, Peters,Regts '19]

  BUT: For $\delta, \epsilon > 0$, $\Delta \geq 3$, exists $C = C(\delta)$,
  for $\lambda < (1-\delta)\lambda_c$, running time $(n/\epsilon)^{C \log \Delta}$.

- All $\Delta \geq 3$, all $\lambda > \lambda_c(\Delta)$: No poly-time to approx. Z for Δ-regular, triangle-free G, unless $NP = RP$

  [Sly '10,Galanis,Stefankovic,V '13, Sly,Sun '13, GSV '15]

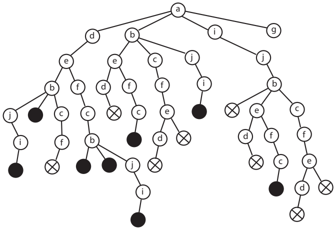
What happens at $\lambda_c(\Delta)$?

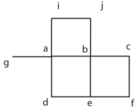
Statistical physics phase transition on infinite Δ-regular tree

Computational phase transition on general max deg. Δ graphs

High-level idea of FPTAS's

[Weitz '06]: For $G = (V, E)$ and vertex $a \in V$, consider T_{saw}:





$$\mathbf{Pr}_{\sigma \sim \mu_T}(\text{root} \notin \sigma) \quad = \quad \mathbf{Pr}_{\sigma \sim \mu_G}(v \notin \sigma)$$

[Barvinok '14]: Consider $Z(\lambda)$ for complex λ.
Suppose $Z(x) \neq 0$ for all x in an open disk D around interval $[0, \lambda]$.
Look at Taylor of $f(x) = \log Z(x)$, then: $f(\lambda) = \sum_{j=0}^{\infty} \frac{\lambda^j}{j!} f^{(j)}(0)$
and $O(\log(n))$ terms gives good approx.

Poly-time for constant Δ: [Patel,Regts '17]
No complex zeros: [Peters,Regts '19]

Glauber dynamics (X_t) = Gibbs Sampler

$X_t \to X_{t+1}$ is defined as follows:

Glauber dynamics (X_t) = Gibbs Sampler

$X_t \rightarrow X_{t+1}$ is defined as follows:

1. Choose v uniformly at random from V.

Glauber dynamics (X_t) = Gibbs Sampler

$X_t \to X_{t+1}$ is defined as follows:

1. Choose v uniformly at random from V.

$$X' = \begin{cases} X_t \cup \{v\} & \text{with probability } \lambda/(1+\lambda) \\ X_t \setminus \{v\} & \text{with probability } 1/(1+\lambda) \end{cases}$$

Glauber dynamics (X_t) = Gibbs Sampler

$X_t \to X_{t+1}$ is defined as follows:

1. Choose v uniformly at random from V.

$$X' = \begin{cases} X_t \cup \{v\} & \text{with probability } \lambda/(1+\lambda) \\ X_t \setminus \{v\} & \text{with probability } 1/(1+\lambda) \end{cases}$$

2. If X' is independent set, then $X_{t+1} = X'$, otherwise $X_{t+1} = X_t$

Stationary distribution is Gibbs distribution: $\mu(X) = \frac{\lambda^{|X|}}{Z}$

Glauber dynamics (X_t) = Gibbs Sampler

$X_t \to X_{t+1}$ is defined as follows:

1. Choose v uniformly at random from V.

$$X' = \begin{cases} X_t \cup \{v\} & \text{with probability } \lambda/(1+\lambda) \\ X_t \setminus \{v\} & \text{with probability } 1/(1+\lambda) \end{cases}$$

2. If X' is independent set, then $X_{t+1} = X'$, otherwise $X_{t+1} = X_t$

Stationary distribution is Gibbs distribution: $\mu(X) = \frac{\lambda^{|X|}}{Z}$

Mixing Time: $T_{mix} := \min\{t : \text{ for all } X_0,\ d_{tv}(X_t, \mu) \leq 1/4\}$

Then $T_{\rm mix}(\epsilon) \leq \log(1/\epsilon)\, T_{\rm mix}$.

Glauber dynamics (X_t) = Gibbs Sampler

$X_t \to X_{t+1}$ is defined as follows:

1. Choose v uniformly at random from V.

$$X' = \begin{cases} X_t \cup \{v\} & \text{with probability } \lambda/(1+\lambda) \\ X_t \setminus \{v\} & \text{with probability } 1/(1+\lambda) \end{cases}$$

2. If X' is independent set, then $X_{t+1} = X'$, otherwise $X_{t+1} = X_t$

Stationary distribution is Gibbs distribution: $\mu(X) = \frac{\lambda^{|X|}}{Z}$

Mixing Time: $T_{mix} := \min\{t : \text{ for all } X_0,\ d_{tv}(X_t, \mu) \leq 1/4\}$

Then $T_{\mathrm{mix}}(\epsilon) \leq \log(1/\epsilon) T_{\mathrm{mix}}$.

Recall, $d_{\mathrm{TV}}(\mu, \nu) = \frac{1}{2} \sum_{\sigma \in \Omega} |\mu(\sigma) - \nu(\sigma)|$.

Our Results

Our Results

Theorem

For all $\delta > 0$*, there exists* $\Delta_0 = \Delta_0(\delta)$*:*

all $G = (V, E)$ *of max degree* $\Delta \geq \Delta_0$ *and girth* ≥ 7*,*

all $\lambda < (1 - \delta)\lambda_c(\Delta)$*,*

$$T_{\mathrm{mix}} = O\left(n \log n\right).$$

Our Results

Theorem

For all $\delta > 0$, there exists $\Delta_0 = \Delta_0(\delta)$:

all $G = (V, E)$ of max degree $\Delta \geq \Delta_0$ and girth ≥ 7,

all $\lambda < (1 - \delta)\lambda_c(\Delta)$,

$$T_{\mathrm{mix}} = O\left(n \log n\right).$$

Corollaries

- An $O^*(n^2)$ FPRAS for estimating the partition function Z.
- $T_{mix} = O(n \log n)$ when $\lambda \leq (1 - \delta)\lambda_c(\Delta)$ for:
  - random Δ-regular graphs
  - random Δ-regular bipartite graphs

Coupling of Markov Chains

Consider a Markov chain (Ω, P).

Coupling is a joint process $\omega = (X_t, Y_t)$ on $\Omega \times \Omega$ where:

$$X_t \sim P \text{ and } Y_t \sim P$$

More precisely, for all $A, B, C \in \Omega$,

$$\begin{aligned} \mathbf{Pr}\left(X_{t+1} = C \mid X_t = A, Y_t = B\right) &= P(A, C) \\ \mathbf{Pr}\left(X_{t+1} = C \mid X_t = A, Y_t = B\right) &= P(B, C) \end{aligned}$$

Coupling of Markov Chains

Consider a Markov chain (Ω, P).

Coupling is a joint process $\omega = (X_t, Y_t)$ on $\Omega \times \Omega$ where:

$$X_t \sim P \text{ and } Y_t \sim P$$

More precisely, for all $A, B, C \in \Omega$,

$$\begin{aligned}
\mathbf{Pr}(X_{t+1} = C \mid X_t = A, Y_t = B) &= P(A, C) \\
\mathbf{Pr}(X_{t+1} = C \mid X_t = A, Y_t = B) &= P(B, C)
\end{aligned}$$

Intuition:

$(X_t \to X_{t+1}) \sim P$ and $(Y_t \to Y_{t+1}) \sim P$ can correlate by ω.

Let X_0 be arbitrary, and $Y_0 \sim \pi$. Once $X_T = Y_T$ then $X_T \sim \pi$.

Coupling of Markov Chains

Consider a Markov chain (Ω, P).

Coupling is a joint process $\omega = (X_t, Y_t)$ on $\Omega \times \Omega$ where:

$$X_t \sim P \text{ and } Y_t \sim P$$

More precisely, for all $A, B, C \in \Omega$,

$$\begin{aligned} \mathbf{Pr}\left(X_{t+1} = C \mid X_t = A, Y_t = B\right) &= P(A, C) \\ \mathbf{Pr}\left(X_{t+1} = C \mid X_t = A, Y_t = B\right) &= P(B, C) \end{aligned}$$

Intuition:

$(X_t \rightarrow X_{t+1}) \sim P$ and $(Y_t \rightarrow Y_{t+1}) \sim P$ can correlate by ω.

Let X_0 be arbitrary, and $Y_0 \sim \pi$. Once $X_T = Y_T$ then $X_T \sim \pi$.

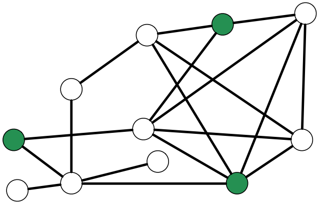
Coupling time:

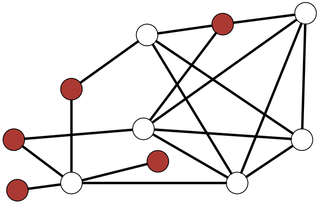
$$T_{\text{couple}} = \max_{A,B \in \Omega} \min\{t : \mathbf{Pr}\left(X_t \neq Y_t \mid X_0 = A, Y_0 = B\right) \leq 1/4.\}$$

$$T_{\text{mix}} \leq T_{\text{couple}}$$

Coupling for Independent Sets

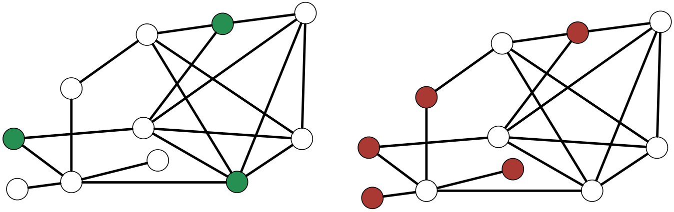
Consider a pair of independent sets X_t and Y_t:



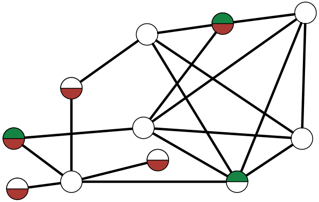


Coupling for Independent Sets

Consider a pair of independent sets X_t and Y_t:

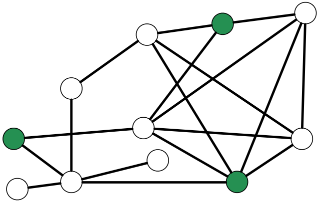


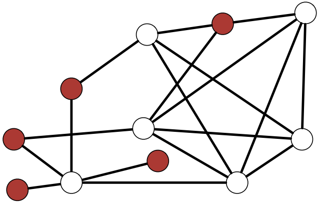
Look at $\frac{X_t}{Y_t}$:



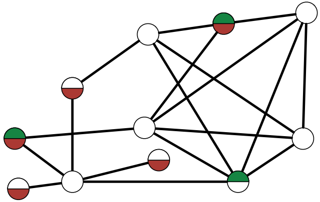
Coupling for Independent Sets

Consider a pair of independent sets X_t and Y_t:





Look at $\frac{X_t}{Y_t}$:

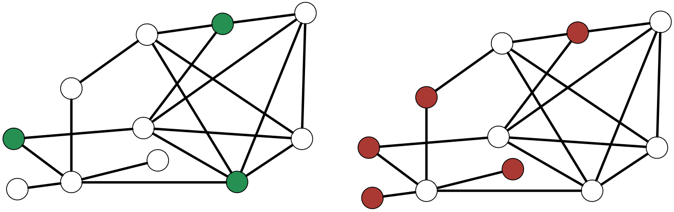


Identity Coupling:

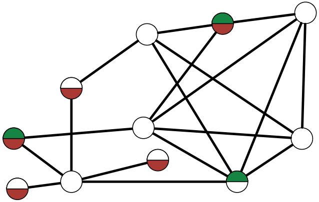
Update same v_t, attempt to add to both or remove from both.

Coupling for Independent Sets

Consider a pair of independent sets X_t and Y_t:



Look at $\frac{X_t}{Y_t}$:



Identity Coupling:

Update same v_t, attempt to add to both or remove from both.

How to analyze???

COUPLING FOR BOUNDING T_{mix}

For all X_t, Y_t, define a coupling: $(X_t, Y_t) \to (X_{t+1}, Y_{t+1})$.

Look at Hamming distance: $H(X_t, Y_t) = |\{v : X_t(v) \neq Y_t(v)\}|$.

If for all X_t, Y_t,

$$\mathbb{E}\left[H(X_{t+1}, Y_{t+1}) \,|\, X_t, Y_t\right] \leq (1 - C/n) H(X_t, Y_t),$$

$$\begin{aligned}
\text{Then,} \quad \mathbf{Pr}\left(A_T \neq B_T\right) &\leq \mathbb{E}\left[H(A_T, B_T)\right] \\
&\leq H(A_0, B_0)(1 - C/n)^T \\
&\leq n \exp(-C/n) \\
&\leq 1/4 \quad \text{for } T = O(n \log n).
\end{aligned}$$

Path coupling: Suffices to consider pairs where $H(X_t, Y_t) = 1$.

Coupling for bounding T_{mix}

For all X_t, Y_t, define a coupling: $(X_t, Y_t) \to (X_{t+1}, Y_{t+1})$.

Look at Hamming distance: $H(X_t, Y_t) = |\{v : X_t(v) \neq Y_t(v)\}|$.

If for all X_t, Y_t,

$$\mathbb{E}\left[H(X_{t+1}, Y_{t+1}) \,|\, X_t, Y_t\right] \leq (1 - C/n) H(X_t, Y_t),$$

$$\begin{aligned}
\text{Then,} \quad \mathbf{Pr}\left(A_T \neq B_T\right) &\leq \mathbb{E}\left[H(A_T, B_T)\right] \\
&\leq H(A_0, B_0)(1 - C/n)^T \\
&\leq n \exp(-C/n) \\
&\leq 1/4 \quad \text{for } T = O(n \log n).
\end{aligned}$$

Path coupling: Suffices to consider pairs where $H(X_t, Y_t) = 1$.

Can replace $H()$:

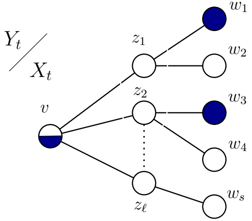
For $\Phi : V \to \mathbb{R}_{\geq 1}$, let $\Phi(X, Y) = \sum_{v \in X \oplus Y} \Phi_v$.

Key: if $X \neq Y$ then $\Phi(X, Y) \geq 1$.

Hence, $\mathbf{Pr}\left(X_t \neq Y_t\right) \leq \mathbb{E}\left[\Phi(X_t, Y_t)\right]$.

Path Coupling with Hamming Distance

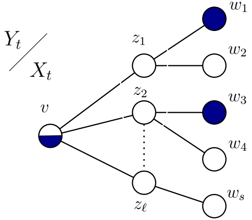
$$\mathbb{E}\left[H(X_{t+1}, Y_{t+1})\right] = H(X_t, Y_t) - \frac{1}{n} + \sum_{z_i} \Pr[z_i \in Y_{t+1}]$$



Coupling: update same vertex, attempt add $\frac{\lambda}{1+\lambda}$, remove $\frac{1}{1+\lambda}$.

Path Coupling with Hamming Distance

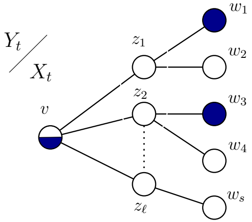
$$\mathbb{E}\left[H(X_{t+1}, Y_{t+1})\right] = H(X_t, Y_t) - \frac{1}{n} + \sum_{z_i} \Pr[z_i \in Y_{t+1}]$$



Coupling: update same vertex, attempt add $\frac{\lambda}{1+\lambda}$, remove $\frac{1}{1+\lambda}$.

Path Coupling with Hamming Distance

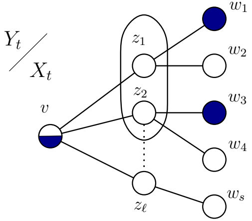
$$\mathbb{E}\left[H(X_{t+1}, Y_{t+1})\right] \;=\; H(X_t, Y_t) - \frac{1}{n} + \sum_{z_i} \Pr[z_i \in Y_{t+1}]$$



Coupling: update same vertex, attempt add $\frac{\lambda}{1+\lambda}$, remove $\frac{1}{1+\lambda}$.

Path Coupling with Hamming Distance

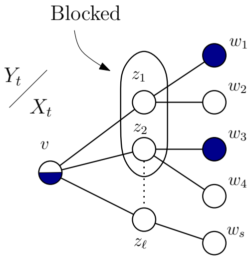
$$\mathbb{E}\left[H(X_{t+1}, Y_{t+1})\right] \quad = \quad H(X_t, Y_t) - \frac{1}{n} + \sum_{z_i} \Pr[z_i \in Y_{t+1}]$$



Coupling: update same vertex, attempt add $\frac{\lambda}{1+\lambda}$, remove $\frac{1}{1+\lambda}$.

Path Coupling with Hamming Distance

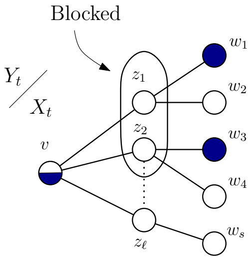
$$\mathbb{E}\left[H(X_{t+1}, Y_{t+1})\right] \quad = \quad H(X_t, Y_t) - \frac{1}{n} + \sum_{z_i} \Pr[z_i \in Y_{t+1}]$$



Coupling: update same vertex, attempt add $\frac{\lambda}{1+\lambda}$, remove $\frac{1}{1+\lambda}$.

Path Coupling with Hamming Distance

$$\begin{aligned}
\mathbb{E}\left[H(X_{t+1}, Y_{t+1})\right] &= H(X_t, Y_t) - \frac{1}{n} + \sum_{z_i} \Pr[z_i \in Y_{t+1}] \\
&= (1 - \frac{1}{n}) + \frac{1}{n} \sum_{z_i} \mathbf{1}\{z_i \text{ unblocked}\} \frac{\lambda}{1+\lambda} \\
&\leq 1 - \frac{1}{n} + \frac{\Delta}{n} \frac{\lambda}{1+\lambda}
\end{aligned}$$

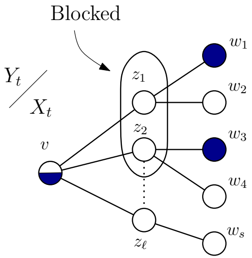


Coupling: update same vertex, attempt add $\frac{\lambda}{1+\lambda}$, remove $\frac{1}{1+\lambda}$.

Path Coupling with Hamming Distance

$$\begin{aligned}
\mathbb{E}\left[H(X_{t+1}, Y_{t+1})\right] &= H(X_t, Y_t) - \frac{1}{n} + \sum_{z_i} \Pr[z_i \in Y_{t+1}] \\
&= (1 - \frac{1}{n}) + \frac{1}{n} \sum_{z_i} \mathbf{1}\{z_i \text{ unblocked}\} \frac{\lambda}{1+\lambda} \\
&\leq 1 - \frac{1}{n} + \frac{\Delta}{n} \frac{\lambda}{1+\lambda} \quad < 1
\end{aligned}$$

Requires: $\lambda < 1/(\Delta - 1)$



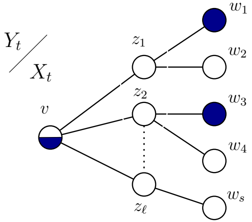
Coupling: update same vertex, attempt add $\frac{\lambda}{1+\lambda}$, remove $\frac{1}{1+\lambda}$.

Path Coupling with Φ

$$\mathbb{E}\left[\Phi(X_{t+1}, Y_{t+1}) \middle| X_t, Y_t\right] = \left(1 - \frac{1}{n}\right)\Phi_v + \sum_{z_i} \Pr[z_i \in Y_{t+1}] \cdot \Phi_{z_i}$$

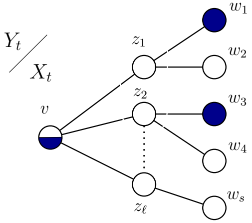
Path Coupling with Φ

$$\mathbb{E}\left[\Phi(X_{t+1}, Y_{t+1}) \middle| X_t, Y_t\right] = \left(1 - \frac{1}{n}\right)\Phi_v + \sum_{z_i} \Pr[z_i \in Y_{t+1}] \cdot \Phi_{z_i}$$



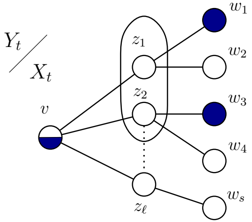
Path Coupling with Φ

$$\mathbb{E}\left[\Phi(X_{t+1}, Y_{t+1}) \middle| X_t, Y_t\right] = \left(1 - \frac{1}{n}\right)\Phi_v + \sum_{z_i} \Pr[z_i \in Y_{t+1}] \cdot \Phi_{z_i}$$



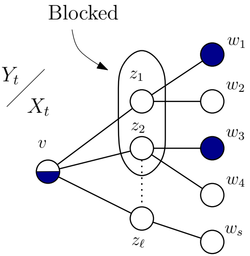
Path Coupling with Φ

$$\mathbb{E}\left[\Phi(X_{t+1}, Y_{t+1}) \middle| X_t, Y_t\right] = \left(1 - \frac{1}{n}\right)\Phi_v + \sum_{z_i} \Pr[z_i \in Y_{t+1}] \cdot \Phi_{z_i}$$



Path Coupling with Φ

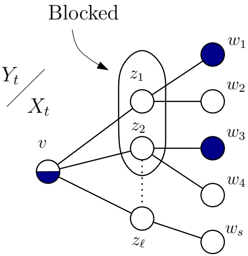
$$\mathbb{E}\left[\Phi(X_{t+1}, Y_{t+1}) \middle| X_t, Y_t\right] = \left(1 - \frac{1}{n}\right)\Phi_v + \sum_{z_i} \Pr[z_i \in Y_{t+1}] \cdot \Phi_{z_i}$$



Path Coupling with Φ

$$\mathbb{E}\left[\Phi(X_{t+1}, Y_{t+1}) \middle| X_t, Y_t\right] = \left(1 - \frac{1}{n}\right)\Phi_v + \sum_{z_i} \Pr[z_i \in Y_{t+1}] \cdot \Phi_{z_i}$$

$$= \left(1 - \frac{1}{n}\right)\Phi_v + \frac{1}{n}\sum_{z_i} \mathbf{1}\{z_i \text{ unblocked}\}\frac{\lambda}{1+\lambda}\Phi_{z_i}$$

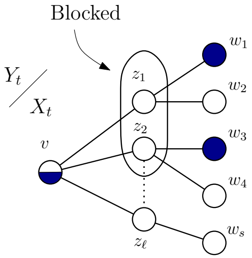


Path Coupling with Φ

$$\mathbb{E}\left[\Phi(X_{t+1}, Y_{t+1}) \middle| X_t, Y_t\right] = \left(1 - \frac{1}{n}\right)\Phi_v + \sum_{z_i} \Pr[z_i \in Y_{t+1}] \cdot \Phi_{z_i}$$

$$= \left(1 - \frac{1}{n}\right)\Phi_v + \frac{1}{n}\sum_{z_i} \mathbf{1}\{z_i \text{ unblocked}\}\frac{\lambda}{1+\lambda}\Phi_{z_i} \quad < \Phi_v$$

Want: $\quad \Phi_v > \frac{\lambda}{1+\lambda}\sum_{z_i} \mathbf{1}\{z_i \text{ unblocked in } Y_t\} \cdot \Phi_{z_i}$



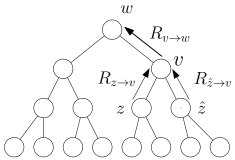
Belief Propagation on trees

For tree T and given λ, compute:

$q(v, w) = \mu(v \text{ occupied} | w \text{ unoccupied})$

$$R_{v\to w} = \frac{q(v,w)}{1 - q(v,w)}$$

$$R_{v\to w} = \lambda \prod_{z\in N(v)\setminus\{w\}} \frac{1}{1 + R_{z\to v}}$$



BP starts from arbitrary $R^0_{v\to w}$,
then iterates:

$$R^i_{v\to w} = \lambda \prod_{z\in N(v)\setminus\{w\}} \frac{1}{1 + R^{i-1}_{z\to v}}$$

BP and Gibbs distribution on trees

Convergence on trees

For $i >$ max-depth, for every initial (R^0):

$$R^i_{v\to w} = R^*_{v\to w}$$

In turn

$$\mu(v \text{ occupied}|w \text{ unoccupied}) = q^* = \frac{R^*_{v\to w}}{1 + R^*_{v\to w}}$$

BP is an elaborate version of *Dynamic Programing*

BP Convergence for girth ≥ 6

Loopy Belief Propagation: Run BP on general $G = (V, E)$. For all $v \in V, w \in N(v)$:

$$R^i_{v \to w} = \lambda \prod_{z \in N(v) \setminus \{w\}} \frac{1}{1 + R^{i-1}_{z \to v}} \quad \text{and} \quad q^i(v, w) = \frac{R^i_{v \to w}}{1 + R^i_{v \to w}}$$

BP Convergence for girth ≥ 6

Loopy Belief Propagation: Run BP on general $G = (V, E)$. For all $v \in V, w \in N(v)$:

$$R^i_{v \to w} = \lambda \prod_{z \in N(v) \setminus \{w\}} \frac{1}{1 + R^{i-1}_{z \to v}} \quad \text{and} \quad q^i(v, w) = \frac{R^i_{v \to w}}{1 + R^i_{v \to w}}$$

Does it converge? If so, to what?

BP Convergence for girth ≥ 6

Loopy Belief Propagation: Run BP on general $G = (V, E)$. For all $v \in V, w \in N(v)$:

$$R^i_{v \to w} = \lambda \prod_{z \in N(v) \setminus \{w\}} \frac{1}{1 + R^{i-1}_{z \to v}} \quad \text{and} \quad q^i(v, w) = \frac{R^i_{v \to w}}{1 + R^i_{v \to w}}$$

Does it converge? If so, to what?

For $\lambda < \lambda_c$: $R()$ has a unique fixed point R^*.

BP Convergence for girth ≥ 6

Loopy Belief Propagation: Run BP on general $G = (V, E)$. For all $v \in V, w \in N(v)$:

$$R^i_{v\to w} = \lambda \prod_{z\in N(v)\setminus\{w\}} \frac{1}{1 + R^{i-1}_{z\to v}} \quad \text{and} \quad q^i(v, w) = \frac{R^i_{v\to w}}{1 + R^i_{v\to w}}$$

Does it converge? If so, to what?

For $\lambda < \lambda_c$: $R()$ has a unique fixed point R^*.

Theorem

Let $\delta, \epsilon > 0$, $\Delta_0 = \Delta_0(\delta, \epsilon)$ and $C = C(\delta, \epsilon)$.
For G of max degree $\Delta \geq \Delta_0$ and girth ≥ 6, all $\lambda < (1-\delta)\lambda_c(\Delta)$:
for $i \geq C$, for all $v \in V$, $w \in N(v)$,

$$\left| \frac{q^i(v, w)}{\mu(v \textit{ is occupied} \mid w \textit{ is unoccupied})} - 1 \right| \leq \epsilon$$

Unblocked Neighbors and loopy BP

Recall, loopy BP estimate that z is unoccupied:

$$R_z^i = \lambda \prod_{y \in N(v)} \frac{1}{1 + R_y^{i-1}}$$

Loopy BP estimate that z is unblocked:

$$\omega_z^i = \prod_{y \in N(z)} \frac{1}{1 + \lambda \cdot \omega_y^{i-1}}$$

For $\lambda < \lambda_c$:

Since R converges to unique fixed point R^*,

thus ω converges to unique fixed point ω^*.

We'll prove (but don't know a priori):

$$\omega^*(z) \approx \mu(z \text{ is unblocked})$$

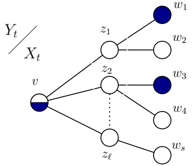
Back to Path Coupling

worst case condition

$$\Phi_v > \frac{\lambda}{1+\lambda} \sum_{z_i} \mathbf{1}\{z_i \text{ unblocked}\} \cdot \Phi_{z_i}$$

when X_t, Y_t "behave" like ω^*

$$\Phi_v > \frac{\lambda}{1+\lambda} \sum_{z_i} \omega^*(z_i) \cdot \Phi_{z_i}$$



Finding Φ

Reformulation

Goal: Find Φ such that

$$\Phi_v > \sum_{z \in N(v)} \frac{\lambda \omega^*(z)}{1 + \lambda \omega^*(z)} \, \Phi_z$$

Finding Φ

Reformulation

Goal: Find Φ such that

$$\Phi_v > \sum_{z \in N(v)} \frac{\lambda \omega^*(z)}{1 + \lambda \omega^*(z)} \Phi_z$$

Define $n \times n$ matrix $\mathcal{C}$

$$\mathcal{C}(v, z) = \begin{cases} \frac{\lambda \omega^*(z)}{1+\lambda \omega^*(z)} & \text{if } z \in N(v) \\ 0 & \text{otherwise} \end{cases}$$

Finding Φ

Reformulation

Goal: Find Φ such that

$$\Phi_v > \sum_{z \in N(v)} \frac{\lambda \omega^*(z)}{1 + \lambda \omega^*(z)} \Phi_z$$

Define $n \times n$ matrix $\mathcal{C}$

$$\mathcal{C}(v, z) = \begin{cases} \frac{\lambda \omega^*(z)}{1+\lambda \omega^*(z)} & \text{if } z \in N(v) \\ 0 & \text{otherwise} \end{cases}$$

Rephrased: Find vector $\Phi \in \mathbb{R}^V_{\geq 1}$ such that

$$\mathcal{C}\ \Phi < \Phi$$

Connections with Loopy BP

Recall, BP operator for unblocked: $F(\omega)(z) = \prod_{y \in N(z)} \frac{1}{1 + \lambda\omega(y)}$

It has Jacobian: $J(v, u) = \left| \frac{\partial F(\omega)(v)}{\partial \omega(u)} \right| = \begin{cases} \frac{\lambda F(\omega)(v)}{1+\lambda\omega(u)} & \text{if } u \in N(v) \\ 0 & \text{otherwise} \end{cases}$

Let $J^* = J|_{\omega=\omega^*}$ denote the Jacobian at the fixed point ω^*.

Key fact: $\mathcal{C} = D^{-1} J^* D$,

where D is diagonal matrix with $D(v, v) = \omega^*(v)$

Connections with Loopy BP

Recall, BP operator for unblocked: $F(\omega)(z) = \prod_{y \in N(z)} \frac{1}{1 + \lambda \omega(y)}$

It has Jacobian: $J(v, u) = \left| \frac{\partial F(\omega)(v)}{\partial \omega(u)} \right| = \begin{cases} \frac{\lambda F(\omega)(v)}{1+\lambda\omega(u)} & \text{if } u \in N(v) \\ 0 & \text{otherwise} \end{cases}$

Let $J^* = J|_{\omega=\omega^*}$ denote the Jacobian at the fixed point ω^*.

Key fact: $\mathcal{C} = D^{-1} J^* D$,

where D is diagonal matrix with $D(v, v) = \omega^*(v)$

Fixed point ω^* is Jacobian attractive so all eigenvalues < 1.

Principal eigenvector Φ is good coupling distance.

Key Results

Proof approach:

- Find good Φ when locally X_t, Y_t "behave" like ω^*
- dynamics gets "*local uniformity*":
  $O(n \log \Delta)$ steps looks locally like ω^*. builds on [Hayes '13]
- Disagreements don't spread too fast

  builds on [Dyer-Frieze-Hayes-V '13]

Outline

Proof approach:

- Find good Φ when locally X_t, Y_t "behave" like ω^*
  –dynamics gets "*local uniformity*" builds on [Hayes '13]

- Disagreements don't spread too fast:
  builds on [Dyer-Frieze-Hayes-V '13]

Outline

Proof approach:

- Find good Φ when locally X_t, Y_t "behave" like ω^*

  –dynamics gets "*local uniformity*" builds on [Hayes '13]

  For any X_0, when $\lambda < \lambda_c$ and girth ≥ 7,
  with prob. $\geq 1 - \exp(-\Omega(\Delta))$, for $t = \Omega(n \log \Delta)$:

  $$\#\{\text{Unblocked Neighbors of } v \text{ in } X_t\} < \sum_{z \in N(v)} \omega^*(z) + \epsilon\Delta.$$

- Disagreements don't spread too fast:

  builds on [Dyer-Frieze-Hayes-V '13]

Outline

Proof approach:

- Find good Φ when locally X_t, Y_t "behave" like ω^*

  –dynamics gets "*local uniformity*" builds on [Hayes '13]

  For any X_0, when $\lambda < \lambda_c$ and girth ≥ 7,
  with prob. $\geq 1 - \exp(-\Omega(\Delta))$, for $t = \Omega(n \log \Delta)$:

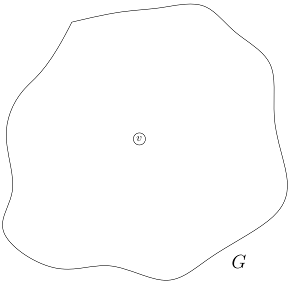
  $$\#\{\text{Unblocked Neighbors of } v \text{ in } X_t\} < \sum_{z \in N(v)} \omega^*(z) + \epsilon\Delta.$$

- Disagreements don't spread too fast:

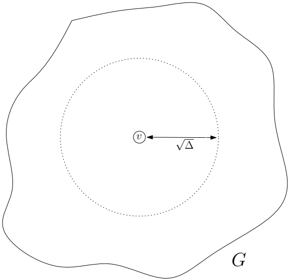
  builds on [Dyer-Frieze-Hayes-V '13]

  For (X_0, Y_0) differ only at v, for $T = O(n \log \Delta)$, $r = O(\sqrt{\Delta})$,

  $$\mathbf{Pr}\left(X_T \oplus Y_T \subset B_r(v)\right) \geq 1 - \exp(\Omega(\sqrt{\Delta}))$$

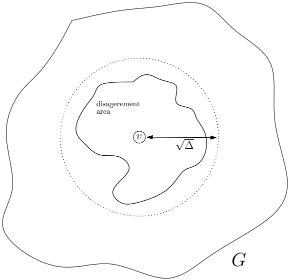


1. Initially: single disagreement at v.



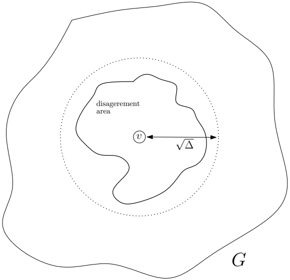
1. Initially: single disagreement at v.
2. Run the chains for $O(n \log \Delta)$ steps: "burn-in".

Rapid Mixing with Uniformity [Dyer-Frieze-Hayes-V '13]



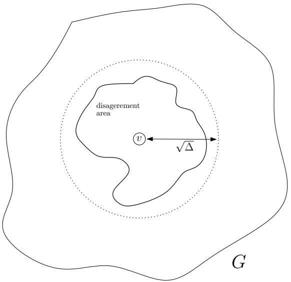
1. Initially: single disagreement at v.
2. Run the chains for $O(n \log \Delta)$ steps: "burn-in".
3. The disagreements might spread during this burn-in.

Rapid Mixing with Uniformity [Dyer-Frieze-Hayes-V '13]



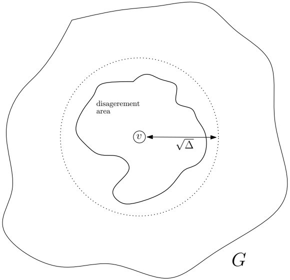
1. Initially: single disagreement at v.
2. Run the chains for $O(n \log \Delta)$ steps: "burn-in".
3. The disagreements might spread during this burn-in.
4. The disagreements do not escape the ball B, whp.

Rapid Mixing with Uniformity [Dyer-Frieze-Hayes-V '13]



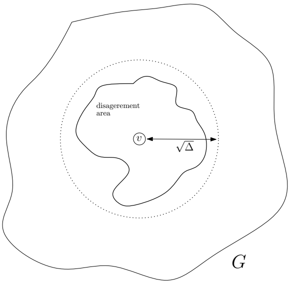
1. Initially: single disagreement at v.
2. Run the chains for $O(n \log \Delta)$ steps: "burn-in".
3. The disagreements might spread during this burn-in.
4. The disagreements do not escape the ball B, whp.
5. The entire ball B has uniformity, whp.

Rapid Mixing with Uniformity [Dyer-Frieze-Hayes-V '13]



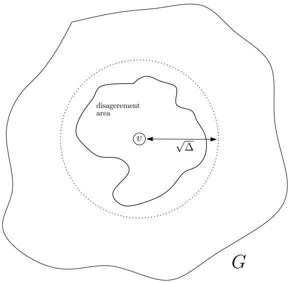
1. Initially: single disagreement at v.
2. Run the chains for $O(n \log \Delta)$ steps: "burn-in".
3. The disagreements might spread during this burn-in.
4. The disagreements do not escape the ball B, whp.
5. The entire ball B has uniformity, whp.
6. Interpolate and do path coupling for the disagree pairs in B, ... pairs have local uniformity

Rapid Mixing with Uniformity [Dyer-Frieze-Hayes-V '13]



1. Initially: single disagreement at v.
2. Run the chains for $O(n \log \Delta)$ steps: "burn-in".
3. The disagreements might spread during this burn-in.
4. The disagreements do not escape the ball B, whp.
5. The entire ball B has uniformity, whp.
6. Interpolate and do path coupling for the disagree pairs in B, ... pairs have local uniformity and Φ gives contraction

Rapid Mixing with Uniformity [Dyer-Frieze-Hayes-V '13]



1. Initially: single disagreement at v.
2. Run the chains for $O(n \log \Delta)$ steps: "burn-in".
3. The disagreements might spread during this burn-in.
4. The disagreements do not escape the ball B, whp.
5. The entire ball B has uniformity, whp.
6. Interpolate and do path coupling for the disagree pairs in B, ... pairs have local uniformity and Φ gives contraction
7. Run $O(n)$ steps to get expected $\#$ of disagreements $< 1/8$.

Questions

What happens at λ_c?

Questions

What happens at λ_c?

THANK YOU!