Convergence of MCMC and Loopy BP in the Tree Uniqueness Region for the Hard-Core Model

Charis Efthymiou
efthymiou@gmail.com

University of Frankfurt

joint work with:
T. Hayes, D. Štefankovič, E. Vigoda, Y. Yin

Workshop on Random Instances and Phase Transitions
Simons Institute 2-6 May 2016
Hard-Core Model

Given a graph $G = (V, E)$ and a fugacity $\lambda > 0$, for each independent set $\sigma$ we have
$$
\mu(\sigma) = \frac{\lambda |\sigma|}{Z},
$$
where $Z = \sum_{\sigma} \lambda |\sigma|$ is the partition function $Z(G, \lambda)$ is the partition function.
Given a graph $G = (V, E)$ and a fugacity $\lambda > 0$, for each independent set $\sigma$ we have

$$
\mu(\sigma) = \frac{\lambda |\sigma|}{Z},
$$

where $Z = \sum_{\sigma} \lambda |\sigma|$ is the partition function.
Gibbs Distribution

Given a graph $G = (V, E)$ and fugacity $\lambda > 0$, for each independent set $\sigma$ we have

$$\mu(\sigma) = \frac{\lambda^{|\sigma|}}{Z},$$

where $Z = Z(G, \lambda)$ is the partition function.
Hard-Core Model

Gibbs Distribution

Given a graph $G = (V, E)$ and fugacity $\lambda > 0$, for each independent set $\sigma$ we have

$$\mu(\sigma) = \frac{\lambda^{\lvert \sigma \rvert}}{Z},$$

where

$$Z = \sum_{\sigma} \lambda^{\lvert \sigma \rvert}$$
Gibbs Distribution

Given a graph $G = (V, E)$ and fugacity $\lambda > 0$, for each independent set $\sigma$ we have

$$\mu(\sigma) = \frac{\lambda^{\left|\sigma\right|}}{Z},$$

where

$$Z = \sum_\sigma \lambda^{\left|\sigma\right|}$$

$Z = Z(G, \lambda)$ is the partition function.
The problem

Given a graph $G = (V, E)$ and fugacity $\lambda > 0$, compute the partition function $Z(G, \lambda) = \sum_{\sigma} \lambda^{|\sigma|}$

computationally hard problem

$\#P$-complete [Valiant 1979]

focus on the approximation algorithms
Given a graph $G = (V, E)$ and fugacity $\lambda > 0$, compute the partition function

$$Z(G, \lambda) = \sum_{\sigma} \lambda |\sigma|$$
The problem

Given a graph \( G = (V, E) \) and fugacity \( \lambda > 0 \), compute the partition function

\[
Z(G, \lambda) = \sum_{\sigma} \lambda^{|\sigma|}
\]

- computationally hard problem
The problem

Given a graph $G = (V, E)$ and fugacity $\lambda > 0$, compute the partition function

$$Z(G, \lambda) = \sum_{\sigma} \lambda |\sigma|$$

- computationally hard problem
  - $\#P$-complete [Valiant 1979]
The problem

Given a graph \( G = (V, E) \) and fugacity \( \lambda > 0 \), compute the partition function

\[
Z(G, \lambda) = \sum_{\sigma} \lambda |\sigma|
\]

- computationally hard problem
  - \#P-complete [Valiant 1979]
- focus on the approximation algorithms
Approximation Algorithms’ Approach
Approach

Compute estimates of the Gibbs distribution
Approach

Compute estimates of the Gibbs distribution

- Deterministic
Approximation Algorithms’ Approach

Approach

Compute estimates of the Gibbs distribution

- Deterministic

- Randomized
Approximation Algorithms’ Approach

Approach

Compute estimates of the Gibbs distribution

- Deterministic
  - Compute numerically (estimations of) the probability of a configuration

- Randomized

Fully Polynomial Time Approximation Scheme (FPTAS)
in time $\text{poly}(n)$ and $\text{poly}(\epsilon^{-1})$

$\hat{Z} \in (1 \pm \epsilon) Z(G, \lambda)$

Fully Polynomial Time Randomized Approximation Scheme (FPRAS)
in time $\text{poly}(n), \text{poly}(\epsilon^{-1})$ and $\text{poly}(\log(\delta^{-1}))$

$\Pr[\hat{Z} \in (1 \pm \epsilon) Z(G, \lambda)] > 1 - \delta$
Approximation Algorithms’ Approach

Approach

Compute estimates of the Gibbs distribution

- Deterministic
  - Compute numerically (estimations of) the probability of a configuration

- Randomized
  - Generate Samples (approximately) Gibbs distributed
Approach

Compute estimates of the Gibbs distribution

- Deterministic
  - Compute numerically (estimations of) the probability of a configuration
  - Fully Polynomial Time Approximation Scheme (FPTAS)

- Randomized
  - Generate Samples (approximately) Gibbs distributed
  - Fully Polynomial Time Randomized Approximation Scheme (FPRAS)
**Approach**

**Compute estimates of the Gibbs distribution**

- **Deterministic**
  - Compute numerically (estimations of) the probability of a configuration
  - Fully Polynomial Time Approximation Scheme (FPTAS)
    - in time \(\text{poly}(n)\) and \(\text{poly}(\epsilon^{-1})\)
    \[
    \hat{Z} \in (1 \pm \epsilon)Z(G, \lambda)
    \]

- **Randomized**
  - Generate Samples (approximately) Gibbs distributed
  - Fully Polynomial Time Randomized Approximation Scheme (FPRAS)
    - in time \(\text{poly}(n), \text{poly}(\epsilon^{-1})\) and \(\text{poly}(\log(\delta^{-1}))\)
    \[
    \Pr[\hat{Z} \in (1 \pm \epsilon)Z(G, \lambda)] > 1 - \delta
    \]
How well can we approximate

Hardness of approximation [Sly 2010]

For triangle-free $\Delta$-regular graphs, where $\Delta \geq 3$, and for all $\lambda > \lambda_c(\Delta)$, it is NP-hard to approximate the partition function within factor $2^{\gamma n}$. 
How well can we approximate
Hardness of approximation [Sly 2010]

For triangle-free $\Delta$-regular graphs, where $\Delta \geq 3$, and for all $\lambda > \lambda_c(\Delta)$, it is NP-hard to approximate the partition function within factor $2^{\gamma n}$.

- Galanis, Ge, Stefankovic, Vigoda, Yang (2011)
- Sly, Sun (2012)
- Galanis, Stefankovic, Vigoda (2012)
How well can we approximate

Hardness of approximation [Sly 2010]
For triangle-free $\Delta$-regular graphs, where $\Delta \geq 3$, and for all $\lambda > \lambda_c(\Delta)$, it is NP-hard to approximate the partition function within factor $2^{\gamma n}$.

What is $\lambda_c(\Delta)$? [Kelly 1985]
Critical point for “uniqueness/non-uniqueness” phase transition of the hard-core model on $\Delta$ regular trees

$$\lambda_c(\Delta) := \frac{(\Delta - 1)^{\Delta-1}}{(\Delta - 2)^\Delta} \sim \frac{e}{\Delta}$$
$\Delta$-regular tree $T$ of height $h$
Gibbs Uniqueness

$\Delta$-regular tree $T$ of height $h$
Gibbs Uniqueness

\[ \Delta\text{-regular tree } T \text{ of height } h \]

Take two *extreme* configurations on \( L(h) \)
Δ-regular tree $T$ of height $h$

Take two *extreme* configurations on $L(h)$
Gibbs Uniqueness

A $\Delta$-regular tree $T$ of height $h$

Take two *extreme* configurations on $L(h)$
Gibbs Uniqueness

\[ \Delta \text{-regular tree } T \text{ of height } h \]

Take two extreme configurations on \( L(h) \)

For every \( \lambda \) consider

\[ \lim_{h \to \infty} \left| \mu(\cdot|L(h) \text{ occupied}) - \mu(\cdot|L(h) \text{ unoccupied}) \right| = \{0\} \]

Unique

\( \delta \)

Non-Unique

C.Efthymiou (Frankfurt)
Rapid Mixing from Loopy BP
Gibbs Uniqueness

\[ T \text{ of height } h \]

Take two extreme configurations on \( L(h) \)

For every \( \lambda \) consider

\[ \lim_{h \to \infty} ||\mu(\cdot | L(h) \text{ occupied}) - \mu(\cdot | L(h) \text{ unoccupied})||_{\{r\}} = \{0\} \]
Gibbs Uniqueness

$\Delta$-regular tree $T$ of height $h$

Take two *extreme* configurations on $L(h)$

For every $\lambda$ consider

$$\lim_{h \to \infty} ||\mu(\cdot|L(h) \text{ occupied}) - \mu(\cdot|L(h) \text{ unoccupied})|| \{r\}$$
Gibbs Uniqueness

$\Delta$-regular tree $T$ of height $h$
Take two extreme configurations on $L(h)$

For every $\lambda$ consider

$$\lim_{h \to \infty} ||\mu(\cdot|L(h) \text{ occupied}) - \mu(\cdot|L(h) \text{ unoccupied})||_r = \begin{cases} 0 \\ \text{Unique} \\ \text{Non-Unique} \end{cases}$$
Gibbs Uniqueness

$\Delta$-regular tree $T$ of height $h$

Take two extreme configurations on $L(h)$

For every $\lambda$ consider

$$\lim_{h \to \infty} \| \mu(\cdot | L(h) \text{ occupied}) - \mu(\cdot | L(h) \text{ unoccupied}) \|_{\{r\}} = \begin{cases} 0 \end{cases}$$
Δ-regular tree $T$ of height $h$
Take two extreme configurations on $L(h)$

For every $\lambda$ consider

$$\lim_{h \to \infty} ||\mu(\cdot | L(h) \text{ occupied}) - \mu(\cdot | L(h) \text{ unoccupied})||_{\{r\}} = \begin{cases} 0 \\ \delta \end{cases}$$
Gibbs Uniqueness

\[ \Delta \text{-regular tree } T \text{ of height } h \]
Take two extreme configurations on \( L(h) \)

For every \( \lambda \) consider

\[ \lim_{h \to \infty} \| \mu(\cdot|L(h) \text{ occupied}) - \mu(\cdot|L(h) \text{ unoccupied}) \|_{\{r\}} = \begin{cases} 0 & \text{Unique} \\ \delta & \text{non-Unique} \end{cases} \]
\[ \Delta \text{-regular tree } T \text{ of height } h \]

Take two \textit{extreme} configurations on \( L(h) \)

For every \( \lambda \) consider

\[
\lim_{h \to \infty} \| \mu(\cdot | L(h) \text{ occupied}) - \mu(\cdot | L(h) \text{ unoccupied}) \| \{r\} = \begin{cases} 
0 & \text{Unique} \\
\delta & \text{non-Unique}
\end{cases}
\]

\[ \lambda < \lambda_c(\Delta) \iff \text{Gibbs measure is Unique} \]
Gibbs Uniqueness

\[ \Delta \text{-regular tree } T \text{ of height } h \]

Take two extreme configurations on \( L(h) \)

For every \( \lambda \) we compare

\[
\lim_{h \to \infty} ||\mu(\cdot|L(h) \text{ occupied}) - \mu(\cdot|L(h) \text{ unoccupied})||_\{r\} = \begin{cases} 0 & \text{Unique} \\ \delta & \text{non-Unique} \end{cases}
\]

\[ \lambda < \lambda_c(\Delta) \iff \text{we have spatial mixing} \]
Deterministic Algorithms

Weitz's approach \cite{Weitz 2006}

Given $G$ and $\lambda < \lambda_c$, uses tree of self avoiding walks, to organize the computations reduces to dynamic programming. The size of computations depends on the size of the tree in the worst case the tree is exponentially large. (strong) spatial mixing allows to "prune" the tree and still be accurate. This step requires $\lambda < \lambda_c (\Delta)$.

**Weitz’s approach [Weitz 2006]**

Given $G$ and $\lambda < \lambda_c$, uses tree of self-avoiding walks, to organize the computations reduces to dynamic programming. The size of computations depends on the size of the tree, in the worst case the tree is exponentially large. (strong) spatial mixing allows to “prune” the tree and still be accurate. This step requires $\lambda < \lambda_c$ ($\Delta$).

L. Li, P. Lu, and Y. Yin (2012), (2013)
Restrepo, Shin, Tetali, Vigoda, and Yang (2013)
A. Sinclair, P. Srivastava, and Y. Yin (2013)
Weitz’s approach [Weitz 2006]

Given $G$ and $\lambda < \lambda_c$,

- uses tree of self avoiding walks, to organize the computations

for the size of computations depends on the size of the tree
in the worst case the tree is exponentially large

(strong) spatial mixing allows to “prune” the tree and still be
accurate.

this step requires $\lambda < \lambda_c$ ($\Delta L$)
Weitz’s approach [Weitz 2006]

Given $G$ and $\lambda < \lambda_c$,

- uses **tree of self avoiding walks**, to organize the computations
  - reduces to dynamic programming.
Weitz’s approach [Weitz 2006]

Given $G$ and $\lambda < \lambda_c$, 

- uses **tree of self avoiding walks**, to organize the computations
  - reduces to dynamic programming.
- the size of computations depends on the size of the tree
Weitz’s approach [Weitz 2006]

Given $G$ and $\lambda < \lambda_c$, 
- uses tree of self avoiding walks, to organize the computations
  - reduces to dynamic programming.
- the size of computations depends on the size of the tree
  - in the worst case the tree is exponentially large

L. Li, P. Lu, and Y. Yin (2012), (2013)
Restrepo, Shin, Tetali, Vigoda, and Yang (2013)
A. Sinclair, P. Srivastava, and Y. Yin (2013)
Weitz’s approach [Weitz 2006]

Given $G$ and $\lambda < \lambda_c$, 

- uses tree of self avoiding walks, to organize the computations
  - reduces to dynamic programming.
- the size of computations depends on the size of the tree
  - in the worst case the tree is exponentially large
- (strong) spatial mixing allows to “prune” the tree and still be accurate.
Weitz’s approach [Weitz 2006]

Given $G$ and $\lambda < \lambda_c$,

- uses **tree of self avoiding walks**, to organize the computations
  - reduces to dynamic programming.
- the size of computations depends on the size of the tree
  - in the worst case the tree is exponentially large
- **(strong) spatial mixing** allows to “prune” the tree and still be accurate.
  - this step requires $\lambda < \lambda_c(\Delta)$
Deterministic Algorithms

Weitz’s approach [Weitz 2006]

Given $G$ and $\lambda < \lambda_c$,
- uses **tree of self avoiding walks**, to organize the computations
  - reduces to dynamic programming.
- the size of computations depends on the size of the tree
  - in the worst case the tree is **exponentially large**
- **(strong) spatial mixing** allows to “prune” the tree and still be accurate.
  - this step requires $\lambda < \lambda_c(\Delta)$

- L. Li, P. Lu, and Y. Yin (2012), (2013)
- Restrepo, Shin, Tetali, Vigoda, and Yang (2013)
Approximation guarantees

For all $\delta > 0$, there exists constant $C = C(\delta) > 0$, for all $\Delta$ all $G$ of maximum degree $\Delta$, all $\lambda < (1 - \delta)\lambda_c(\Delta)$ all $\epsilon > 0$ Weitz’s algorithm returns an estimation $\hat{Z}$ of the partition function $Z(G, \lambda)$ such that

$$(1 - \epsilon)Z(G, \lambda) \leq \hat{Z} \leq (1 + \epsilon)Z(G, \lambda)$$

in time $O((n/\epsilon)^C \log \Delta)$.
Randomized Algorithm

Given $G$ and $\lambda > 0$, set up an ergodic Markov Chain over the independent sets. The equilibrium distribution is the hard-core model with fugacity $\lambda$. The algorithm simulates the Markov chain and outputs the configuration of the chain after "sufficiently many" steps. The output should be close to the equilibrium distribution. It is desirable that the chain mixes "fast".
Given $G$ and $\lambda > 0$, set up an ergodic Markov Chain over the independent sets. The equilibrium distribution is the hard-core model with fugacity $\lambda$. The algorithm simulates the Markov chain and outputs the configuration of the chain after "sufficiently many" steps. The output should be close to the equilibrium distribution. It is desirable that the chain mixes "fast".
Randomized Algorithm

Markov Chain Monte Carlo

Given $G$ and $\lambda > 0$,
Randomized Algorithm

Markov Chain Monte Carlo

Given $G$ and $\lambda > 0$,

- set up an **ergodic** Markov Chain over the independent sets
Randomized Algorithm

**Markov Chain Monte Carlo**

Given $G$ and $\lambda > 0$,

- set up an **ergodic** Markov Chain over the independent sets
- the **equilibrium distribution** is the hard-core model with fugacity $\lambda$
Randomized Algorithm

Markov Chain Monte Carlo

Given $G$ and $\lambda > 0$,

- set up an **ergodic** Markov Chain over the independent sets
- the **equilibrium distribution** is the hard-core model with fugacity $\lambda$
- the algorithm simulates the Markov chain
Markov Chain Monte Carlo

Given $G$ and $\lambda > 0$, 

- set up an **ergodic** Markov Chain over the independent sets
- the **equilibrium distribution** is the hard-core model with fugacity $\lambda$
- the algorithm simulates the Markov chain
- outputs the configuration of the chain after “sufficiently many” steps
Randomized Algorithm

Markov Chain Monte Carlo

Given $G$ and $\lambda > 0$,

- set up an \textit{ergodic} Markov Chain over the independent sets
- the \textit{equilibrium distribution} is the hard-core model with fugacity $\lambda$
- the algorithm simulates the Markov chain
- outputs the configuration of the chain after “sufficiently many” steps

the output should be close to the equilibrium distribution
Randomized Algorithm

Markov Chain Monte Carlo

Given $G$ and $\lambda > 0$,

- set up an **ergodic** Markov Chain over the independent sets
- the **equilibrium distribution** is the hard-core model with fugacity $\lambda$
- the algorithm simulates the Markov chain
- outputs the configuration of the chain after “sufficiently many” steps

the output should be close to the equilibrium distribution
it is desirable that the chain mixes “fast”
The dynamics

Glauber dynamics ($X_t \rightarrow X_{t+1}$) is defined as follows:

1. Choose $v$ uniformly at random from $V$.

   - $X'_t = \{ X_t \cup \{ v \} \}$ with probability $\frac{\lambda}{1 + \lambda}$.
   - $X'_t = \{ v \}$ with probability $\frac{1}{1 + \lambda}$.

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

The chain converges to the hard-core model with fugacity $\lambda$. 

C.Efthymiou (Frankfurt)  
Rapid Mixing from Loopy BP  
10 / 35
Glauber dynamics \((X_t)\)
The dynamics

Glauber dynamics ($X_t$)

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

1. Choose $v$ uniformly at random from $V$.

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

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

The chain converges to the hard-core model with fugacity $\lambda$. 
The dynamics

Glauber dynamics ($X_t$)

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

1. Choose $v$ uniformly at random from $V$.

The chain converges to the hard-core model with fugacity $\lambda$. 

C.Efthymiou (Frankfurt)  Rapid Mixing from Loopy BP
Glauber dynamics ($X_t$)

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

1. Choose $v$ uniformly at random from $V$.

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

The chain converges to the hard-core model with fugacity $\lambda$. 

C.Efthymiou (Frankfurt)
Rapid Mixing from Loopy BP
The dynamics

Glauber dynamics ($X_t$)

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

1. Choose $v$ uniformly at random from $V$.

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

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

The chain converges to the hard-core model with fugacity $\lambda$. 

C.Efthymiou (Frankfurt)
The dynamics

### Glauber dynamics ($X_t$)

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

1. Choose $v$ uniformly at random from $V$.

\[
X' = \begin{cases} 
X_t \cup \{v\} & \text{with probability } \frac{\lambda}{1 + \lambda} \\
X_t \setminus \{v\} & \text{with probability } \frac{1}{1 + \lambda}
\end{cases}
\]

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

The chain converges to the hard-core model with fugacity $\lambda$. 

Our Results

Theorem

For all $\delta > 0$, there exists $\Delta_0 = \Delta_0(\delta)$ for all graphs $G = (V, E)$ of maximum degree $\Delta \geq \Delta_0$ and girth $\geq 7$, all $\lambda < (1 - \delta) \lambda_c(\Delta)$, the mixing time of the Glauber dynamics satisfies $T_{mix} = O(n \log(n))$. 
Our Results

**Theorem**

For all $\delta > 0$, there exists $\Delta_0 = \Delta_0(\delta)$ for all graphs $G = (V, E)$ of maximum degree $\Delta \geq \Delta_0$ and girth $\geq 7$, all $\lambda < (1 - \delta)\lambda_c(\Delta)$, the mixing time of the Glauber dynamics satisfies

$$T_{mix} = O(n \log(n)).$$
Our Results

**Theorem**

For all $\delta > 0$, there exists $\Delta_0 = \Delta_0(\delta)$ for all graphs $G = (V, E)$ of maximum degree $\Delta \geq \Delta_0$ and girth $\geq 7$, all $\lambda < (1 - \delta)\lambda_c(\Delta)$, the mixing time of the Glauber dynamics satisfies

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

**Mixing Time**

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

Theorem

For all $\delta > 0$, there exists $\Delta_0 = \Delta_0(\delta)$ for all graphs $G = (V, E)$ of maximum degree $\Delta \geq \Delta_0$ and girth $\geq 7$, all $\lambda < (1 - \delta)\lambda_c(\Delta)$, the mixing time of the Glauber dynamics satisfies

$$T_{\text{mix}} = O(n \log(n))$$.

Corollary

The above sampling result yields an FPRAS for estimating the partition function $Z$. The running time is $O^*(n^2)$. 
Our Results

Theorem

For all $\delta > 0$, there exists $\Delta_0 = \Delta_0(\delta)$ for all graphs $G = (V, E)$ of maximum degree $\Delta \geq \Delta_0$ and girth $\geq 7$, all $\lambda < (1 - \delta)\lambda_c(\Delta)$, the mixing time of the Glauber dynamics satisfies

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

Previous work

$T_{mix} = O(n \log(n))$ for Glauber dynamics on $G$ of maximum degree $\Delta$ and $\lambda < 2/(\Delta - 2)$

- Dyer Greenhill, Luby, Vigoda
$O(n \log n)$ mixing for Random Graphs

Corollary

$T_{mix} = O(n \log n)$ for Glauber dynamics with $\lambda \leq (1 - \delta)\lambda_{c}(\Delta)$ for random $\Delta$-regular graph and random $\Delta$-regular bipartite graph.

Mossel, Weitz, Wormald (2009)
Relaxation for girth

“# short cycles in the neighborhood of its vertex in G are not too many”
Relaxation for girth

“# short cycles in the neighborhood of its vertex in $G$ are not too many”

Corollary

$T_{\text{mix}} = O(n \log n)$ for Glauber dynamics with $\lambda \leq (1 - \delta)\lambda_c(\Delta)$ for

Mossel, Weitz, Wormald (2009)
$O(n \log n)$ mixing for Random Graphs

Relaxation for girth

“\# short cycles in the neighborhood of its vertex in $G$ are not too many”

Corollary

$T_{mix} = O(n \log n)$ for Glauber dynamics with $\lambda \leq (1 - \delta)\lambda_c(\Delta)$ for

- random $\Delta$-regular graph
Relaxation for girth

“\# short cycles in the neighborhood of its vertex in $G$ are not too many”

Corollary

$T_{mix} = O(n \log n)$ for Glauber dynamics with $\lambda \leq (1 - \delta)\lambda_c(\Delta)$ for
- random $\Delta$-regular graph
- random $\Delta$-regular bipartite graph
Relaxation for girth

“# short cycles in the neighborhood of its vertex in $G$ are not too many”

Corollary

$T_{mix} = O(n \log n)$ for Glauber dynamics with $\lambda \leq (1 - \delta) \lambda_c(\Delta)$ for

- random $\Delta$-regular graph
- random $\Delta$-regular bipartite graph

Mossel, Weitz, Wormald (2009)
Belief Propagation on trees
Belief Propagation on trees

For $T$ and $\lambda$ compute

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

For every $i \geq 1$

$R^i_{v \rightarrow p} = \lambda \prod_{w \in N(v)} \{p(v)\} \frac{1}{1 + R^{i-1}_{w \rightarrow v}}$
Belief Propagation on trees

For $T$ and $\lambda$ compute

$\mu(v \text{ occupied}|w \text{ unoccupied})$
Belief Propagation on trees

For $T$ and $\lambda$ compute

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

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

For every $i \geq 1$

$$R_i v \rightarrow p(v) = \lambda \prod_{w \in N(v)} \{p(v)\} \frac{1}{1 + R_{i-1} w \rightarrow v}$$
Belief Propagation on trees

For $T$ and $\lambda$ compute

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

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

For every $i \geq 1$

$R_i v \rightarrow p(v) = \lambda \prod_{w \in N(v)} \{p(v)\} \frac{1}{1 + R_{i-1} w \rightarrow v}$
Belief Propagation on trees

For $T$ and $\lambda$ compute

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

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

$$R_{v \rightarrow w} = \frac{q_w(v)}{1 - q_w(v)}$$
Belief Propagation on trees

For $T$ and $\lambda$ compute

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

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

$$R_{v \rightarrow w} = \frac{q_w(v)}{1 - q_w(v)}$$
Belief Propagation on trees

For $T$ and $\lambda$ compute

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

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

$R_{v \rightarrow w} = \frac{q_w(v)}{1 - q_w(v)}$

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

C.Efthymiou (Frankfurt)

Rapid Mixing from Loopy BP
Belief Propagation on trees

For $T$ and $\lambda$ compute

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

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

$$R_{v \rightarrow w} = \frac{q_w(v)}{1 - q_w(v)}$$

$$R_{v \rightarrow w} = \lambda \prod_{z \in N(v) \backslash \{w\}} \frac{1}{1 + R_{z \rightarrow v}}$$
Belief Propagation on trees

For $T$ and $\lambda$ compute

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

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

$$R_{v \rightarrow w} = \frac{q_w(v)}{1 - q_w(v)}$$

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

BP starts from arbitrary $R_{v \rightarrow w}^0$, iterates like

$$R_{v \rightarrow w}^i = \lambda \prod_{z \in N(v) \setminus \{w\}} \frac{1}{1 + R_{z \rightarrow v}^{i-1}}$$
Convergence

There exists $i_0$ such that for every $i \geq i_0$ and every $(v \rightarrow w) \in E$ we have $R_i v \rightarrow w = R^* v \rightarrow w$.

In turn $\mu(v \text{ occupied} | w \text{ unoccupied}) = q^* = R^* v \rightarrow w$. BP is an elaborate use of Dynamic Programming to compute marginal.
Convergence on trees

There exists $i_0$ such that for every $i \geq i_0$ and every $(R^0_v \rightarrow w) \{v,w\} \in E$ we have

$$R^i_v \rightarrow w = R^*_v \rightarrow w$$

In turn

$$\mu(\text{v occupied}|\text{w unoccupied}) = q^* = \frac{R^*_v \rightarrow w}{1 + R^*_v \rightarrow w}$$
Convergence on trees

There exists $i_0$ such that for every $i \geq i_0$ and every $(R^0_{v \rightarrow w})_{\{v,w\} \in E}$ we have

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

In turn

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

BP is an elaborate use of *Dynamic Programming* to compute marginal.
We do not know whether it converges... if does, we do not know where exactly it converges.
We do not know whether it converges
We do not know whether it converges

... if does, we do not know where exactly it converges
Theorem

For $G = (V,E)$ of maximum degree $\Delta \geq \Delta_0$ and girth $\geq 6$, all $\lambda < (1 - \delta) \lambda_c(\Delta)$, the following holds: for $i \geq C$, for all $v \in V, w \in N(v)$,

$$\left| \mu(v \text{ is occupied} \mid w \text{ is unoccupied}) - 1 \right| \leq \epsilon$$

we also have convergence for the BP estimate of $\mu(v \text{ is occupied})$. 
Theorem

For $G = (V, E)$ of maximum degree $\Delta \geq \Delta_0$ and girth $\geq 6$, all $\lambda < (1 - \delta)c(\Delta)$, the following holds: for $i \geq C$, for all $v \in V$, $w \in N(v) \setminus \{w\}$,

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

we also have convergence for the BP estimate of $\mu(v \text{ is occupied} \mid w \text{ is unoccupied})$.
BP Convergence for girth $\geq 6$

\[ R^i_{v \rightarrow w} = \lambda \prod_{z \in N(v) \setminus \{w\}} \frac{1}{1 + R^{i-1}_{z \rightarrow v}} \quad \text{and} \quad q^i_w(v) = \frac{R^i_{v \rightarrow w}}{1 + R^i_{v \rightarrow w}} \]
BP Convergence for girth $\geq 6$

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

Theorem

For $G = (V, E)$ of maximum degree $\Delta \geq \Delta_0$ and girth $\geq 6$, all $\lambda < (1 - \delta)\lambda_c(\Delta)$, the following holds: for $i \geq C$, for all $v \in V$, $w \in N(v)$,

$$\left| \frac{q^i_w(v)}{\mu(v \ is \ occupied \mid w \ is \ unoccupied)} - 1 \right| \leq \epsilon$$
BP Convergence for girth $\geq 6$

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

**Theorem**

For $G = (V, E)$ of maximum degree $\Delta \geq \Delta_0$ and girth $\geq 6$, all $\lambda < (1 - \delta)\lambda_c(\Delta)$, the following holds: for $i \geq C$, for all $v \in V$, $w \in N(v)$,

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

we also have convergence for the BP estimate of $\mu(v \text{ is occupied})$
Consider copies \((X_t, Y_t)\) such that:

\[
X_t \oplus Y_t = \{ v \}
\]

\[
\Phi(X_t+1, Y_t+1) | X_t, Y_t \leq (1 - \gamma) \Phi(X_t, Y_t)
\]

\(\Phi: \Omega \times \Omega \rightarrow \mathbb{R} \geq 1\) is a "distance metric" 

\[
\Phi(X, Y) = \sum_{u \in X \oplus Y} \Phi(u)
\]
Path Coupling [Bubley and Dyer 1997]

Consider copies \((X_s), (Y_s)\) such that \(X_t \oplus Y_t = \{v\}\)
Path Coupling [Bubley and Dyer 1997]

Consider copies \((X_s), (Y_s)\) such that \(X_t \oplus Y_t = \{v\}\)
Path Coupling [Bubley and Dyer 1997]

Consider copies \((X_s), (Y_s)\) such that \(X_t \oplus Y_t = \{v\}\)

\[
\mathbb{E} [\Phi(X_{t+1}, Y_{t+1}) | X_t, Y_t] \leq (1 - \gamma) \Phi(X_t, Y_t).
\]
Path Coupling [Bubley and Dyer 1997]

Consider copies \((X_s, Y_s)\) such that \(X_t \oplus Y_t = \{v\}\)

\[
\mathbb{E} [\Phi(X_{t+1}, Y_{t+1}) | X_t, Y_t] \leq (1 - \gamma) \Phi(X_t, Y_t).
\]

\(\Phi : \Omega \times \Omega \rightarrow \mathbb{R}_{\geq 1}\) is a “distance metric”
Path Coupling [Bubley and Dyer 1997]

Consider copies \((X_s), (Y_s)\) such that \(X_t \oplus Y_t = \{v\}\)

\[
\mathbb{E}[\Phi(X_{t+1}, Y_{t+1}) | X_t, Y_t] \leq (1 - \gamma) \Phi(X_t, Y_t).
\]

\(\Phi: \Omega \times \Omega \to \mathbb{R}_{\geq 1}\) is a “distance metric”

\[
\Phi(X, Y) = \sum_{u \in X \oplus Y} \Phi(u)
\]
Key Results
We don’t know a $\Phi$ that gives contraction for worst-case $X_t, Y_t$. 
Key Results

- We don’t know a $\Phi$ that gives contraction for worst-case $X_t, Y_t$.
- can find $\Phi$ when locally $X_t, Y_t$ “behave” like $R^*$
Key Results

- We don’t know a $\Phi$ that gives contraction for worst-case $X_t, Y_t$.
- can find $\Phi$ when locally $X_t, Y_t$ “behave” like $R^*$
- Glauber dynamics converges locally to $R^*$
Key Results

- We don’t know a $\Phi$ that gives contraction for worst-case $X_t, Y_t$.
- can find $\Phi$ when locally $X_t, Y_t$ “behave” like $R^*$
- Glauber dynamics converges locally to $R^*$
- Given $\Phi$ and convergence of Glauber dynamics we show rapid mixing
Path Coupling Example
Path Coupling Example

\[ Y_t \quad \quad X_t \]

\[ v \quad \quad z_1 \quad w_1 \]
\[ \quad \quad z_2 \quad w_2 \]
\[ \quad \quad \vdots \quad \vdots \]
\[ \quad \quad z_\ell \quad w_\ell \]

\[ \quad \quad w_3 \]
\[ \quad \quad w_4 \]
\[ \quad \quad w_s \]
Path Coupling Example

Expected distance

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

Expected distance

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

Expected distance

\[
E[\Phi(X_{t+1}, Y_{t+1})|X_t, Y_t] = \left(1 - \frac{1}{n}\right)\Phi(v) + \frac{1}{n} \sum_{z_i} \Pr[z_i \in Y_{t+1}] \cdot \Phi(z_i)
\]
Path Coupling Example

Expected distance

\[ \mathbb{E} \left[ \Phi(X_{t+1}, Y_{t+1}) \mid X_t, Y_t \right] = \left( 1 - \frac{1}{n} \right) \Phi(v) + \frac{1}{n} \sum_{z_i \text{ unblocked}} \Phi(z_i) \left( 1 + \frac{\lambda \Phi(z_i)}{1 + \lambda} \right) \]
Path Coupling Example

Path coupling condition

\[ \Phi(v) > \frac{\lambda}{1 + \lambda} \sum_{z_i} \mathbf{1}\{z_i \text{ unblocked in } Y_t\} \cdot \Phi(z_i) \]
Key Results

- We don’t know a $\Phi$ that gives contraction for worst-case $X_t, Y_t$.
- can find $\Phi$ when locally $X_t, Y_t$ “behave” like $R^*$
- Glauber dynamics converges locally to $R^*$
- Given $\Phi$ and convergence of Glauber dynamics we show rapid mixing
Unblocked Neighbors and loopy BP

$$\omega_i(z) = \prod_{y \sim z} 1 + \lambda \cdot \omega_i(y)$$

is the loopy BP estimate of $$z$$ to be unblocked. It converges to a unique fixed point

$$\omega^* = \mu(z \text{ is unblocked})$$
Unblocked Neighbors and loopy BP

\[ \omega^i_z = \prod_{y \sim z} \frac{1}{1 + \lambda \cdot \omega^{i-1}_y} \]
Unblocked Neighbors and loopy BP

\[ \omega^i_z = \prod_{y \sim z} \frac{1}{1 + \lambda \cdot \omega^{i-1}_y} \]

- \( \omega^i(z) \) is the loopy BP estimate of \( z \) to be unblocked
$\omega^i_z = \prod_{y \sim z} \frac{1}{1 + \lambda \cdot \omega^{i-1}_y}$

- $\omega^i(z)$ is the loopy BP estimate of $z$ to be unblocked
- converges to a unique fixed point $\omega^*$
\[
\omega^i_z = \prod_{y \sim z} \frac{1}{1 + \lambda \cdot \omega^{i-1}_y}
\]

- \(\omega^i(z)\) is the loopy BP estimate of \(z\) to be unblocked
- converges to a unique fixed point \(\omega^*\)
- \(\omega^*(z) \approx \mu(z\ \text{is unblocked})\)
Back to Path Coupling
Back to Path Coupling

\[ \Phi(v) > \lambda_1 + \lambda \sum z_i \{z_i \text{unblocked}\} \cdot \Phi(z_i) \]

when \( X_t, Y_t \) "behave" like \( \omega^* \Phi(v) > \lambda_1 + \lambda \sum z_i \omega^*(z_i) \cdot \Phi(z_i) \).
worst case condition

\[ \Phi(v) > \frac{\lambda}{1 + \lambda} \sum_{z_i} 1\{z_i \text{ unblocked}\} \cdot \Phi(z_i) \]
worst case condition

\[ \Phi(v) > \frac{\lambda}{1 + \lambda} \sum_{z_i} 1\{z_i \text{ unblocked}\} \cdot \Phi(z_i) \]

when \( X_t, Y_t \) “behave” like \( \omega^* \)

\[ \Phi(v) > \frac{\lambda}{1 + \lambda} \sum_{z_i} \omega^*(z_i) \cdot \Phi(z_i) \]
Finding \( \Phi \)

For \( \rho = 1 - \delta \), there is \( \Phi \) such that

\[
\rho \cdot \Phi(v) \geq \sum z_i \lambda \omega^* (z_i) + \lambda \omega^* (z_i) \cdot \Phi(z_i)
\]

\( n \times n \) matrix

\[
C(v, z) = \begin{cases} 
\lambda \omega^* (z) + \lambda \omega^* (z) & \text{if } z \notin N(v) \\
0 & \text{otherwise}
\end{cases}
\]

There is a vector \( \Phi \in \mathbb{R}^{V \geq 1} \) such that

\[
C \Phi \leq \rho \cdot \Phi.
\]
Finding $\Phi$

Reformulation

For $\rho = 1 - \delta$, there is $\Phi$ such that

$$\rho \cdot \Phi(v) \geq \sum z_i \lambda \omega^*(z_i) + \lambda \omega^*(z_i) \cdot \Phi(z_i)$$

$n \times n$ matrix $C(v, z) = \{ \lambda \omega^*(z_i) + \lambda \omega^*(z_i) \mid z_i \in N(v) \} 0$ otherwise

There is a vector $\Phi \in \mathbb{R}^V \geq 1$ such that $C \Phi \leq \rho \cdot \Phi$. 

C.Efthymiou (Frankfurt) Rapid Mixing from Loopy BP 23 / 35
Finding $\Phi$

Reformulation

For $\rho = 1 - \delta$, there is $\Phi$ such that

$$
\rho \cdot \Phi(v) \geq \sum_{z_i} \frac{\lambda \omega^*(z_i)}{1 + \lambda \omega^*(z_i)} \cdot \Phi(z_i)
$$
Finding $\Phi$

Reformulation

For $\rho = 1 - \delta$, there is $\Phi$ such that

$$\rho \cdot \Phi(v) \geq \sum_{z_i} \frac{\lambda \omega^*(z_i)}{1 + \lambda \omega^*(z_i)} \cdot \Phi(z_i)$$

$n \times n$ matrix $C$

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

Reformulation

For $\rho = 1 - \delta$, there is $\Phi$ such that

$$\rho \cdot \Phi(v) \geq \sum_{z_i} \frac{\lambda \omega^*(z_i)}{1 + \lambda \omega^*(z_i)} \cdot \Phi(z_i)$$

$n \times n$ matrix $C$

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

There is a vector $\Phi \in \mathbb{R}^V_{\geq 1}$ such that

$$C\Phi \leq \rho \cdot \Phi.$$
Connections with Loopy BP

Jacobian of Loopy BP

$$\mathbf{F}(\omega z) = \prod_{y \in N}(z) \frac{1}{1 + \lambda \omega_y}.$$  

$\mathbf{J}^* = \mathbf{J} | \omega = \omega^*$ denote the Jacobian of $F$ at the fixed point $\omega = \omega^*$.

$\hat{\mathbf{J}} = \mathbf{D}^{-1} \mathbf{J}^* \mathbf{D},$ where $\mathbf{D}$ is diagonal matrix, with $\mathbf{D}(v, v) = \omega^*(v)$.

Relation to Path Coupling

$\hat{\mathbf{J}} = \mathbf{C}$.
Connections with Loopy BP

Jacobian of Loopy BP

\[ F(\omega) = \prod_{y \in N}(z) \] 

\[ J^* = J|_{\omega = \omega^*} \] denote the Jacobian of \( F \) at the fixed point \( \omega = \omega^* \).

\[ \hat{J} = D^{-1}J^*D, \] where \( D \) is a diagonal matrix, with \( D(v, v) = \omega^*(v) \).

Relation to Path Coupling

\[ \hat{J} = C \]
Connections with Loopy BP

Jacobian of Loopy BP

BP Operator

\[ F(\omega_z) = \prod_{y \in N(z)} \frac{1}{1 + \lambda \omega_y}. \]

\( \hat{J} \) denote the Jacobian of \( F \) at the fixed point \( \omega = \omega^* \).

\( \hat{J} = D^{-1}J \hat{D} \), where \( D \) is diagonal matrix, with \( D(v,v) = \omega^*(v) \).
Connections with Loopy BP

Jacobian of Loopy BP

BP Operator

\[ F(\omega_z) = \prod_{y \in N(z)} \frac{1}{1 + \lambda \omega_y}. \]

\[ J^* = J|_{\omega = \omega^*} \] denote the Jacobian of \( F \) at the fixed point \( \omega = \omega^* \).
Connections with Loopy BP

Jacobian of Loopy BP

BP Operator

\[ F(\omega_z) = \prod_{y \in N(z)} \frac{1}{1 + \lambda \omega_y}. \]

\( J^* = J|_{\omega=\omega^*} \) denote the Jacobian of \( F \) at the fixed point \( \omega = \omega^* \).

\[ \hat{J} = D^{-1} J^* D, \]

where \( D \) is diagonal matrix, with \( D(v, v) = \omega^*(v) \)
Connections with Loopy BP

Jacobian of Loopy BP

BP Operator

\[ F(\omega_z) = \prod_{y \in N(z)} \frac{1}{1 + \lambda \omega_y}. \]

\( J^* = J|_{\omega=\omega^*} \) denote the Jacobian of \( F \) at the fixed point \( \omega = \omega^* \).

\[ \hat{J} = D^{-1} J^* D, \]

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

Relation to Path Coupling

\[ \hat{J} = C \]
Reduction to BP

For $\rho = 1 - \delta$, there is a vector $\Phi \in \mathbb{R}^V$ such that $\hat{J}\Phi \leq \rho \cdot \Phi$.

$\hat{J}$ has the same eigenvalues as the Jacobian of BP at the fixed point.

Spectral radius of BP in uniqueness region

We should expect $\rho(\lambda, \Delta) < 1$, because the fixed point $\omega^*$ is attractive.

$\Phi(v) = \sqrt{1 + \lambda \omega^*(v)} \omega^*(v)$

From Perron-Frobenius

What is $\Phi(v)$?

C.Efthymiou (Frankfurt)
For $\rho = 1 - \delta$, there is a vector $\Phi \in \mathbb{R}^V$ such that

$$\hat{J}\Phi \leq \rho \cdot \Phi$$
For $\rho = 1 - \delta$, there is a vector $\Phi \in \mathbb{R}^V$ such that

$$\hat{J}\Phi \leq \rho \cdot \Phi$$

$\hat{J}$ has the same eigenvalues as the Jacobian of BP at the fixed point.
For $\rho = 1 - \delta$, there is a vector $\Phi \in \mathbb{R}^V$ such that

$$\hat{J}\Phi \leq \rho \cdot \Phi$$

$\hat{J}$ has the same eigenvalues as the Jacobian of BP at the fixed point

We should expect $\rho(\lambda, \Delta) < 1$, because the fixed point $\omega^*$ is attractive
Covergence from loopy BP

Reduction to BP Spectral radius

For $\rho = 1 - \delta$, there is a vector $\Phi \in \mathbb{R}^V$ such that

$$\hat{J}\Phi \leq \rho \cdot \Phi$$

$\hat{J}$ has the same eigenvalues as the Jacobian of BP at the fixed point

Spectral radius of BP in uniqueness region

We should expect $\rho(\lambda, \Delta) < 1$, because the fixed point $\omega^*$ is attractive

- $\Phi > 0$ from Perron-Frobenius
Covergence from loopy BP

Reduction to BP Spectral radius

For $\rho = 1 - \delta$, there is a vector $\Phi \in \mathbb{R}^V$ such that

$$\hat{J}\Phi \leq \rho \cdot \Phi$$

$\hat{J}$ has the same eigenvalues as the Jacobian of BP at the fixed point.

Spectral radius of BP in uniqueness region

We should expect $\rho(\lambda, \Delta) < 1$, because the fixed point $\omega^*$ is attractive

- $\Phi > 0$ from Perron-Frobenius

What is $\Phi$

$$\Phi(v) = \sqrt{\frac{1 + \lambda \omega^*(v)}{\omega^*(v)}}$$
Key Results

- We don’t know a $\Phi$ that gives contraction for worst-case $X_t, Y_t$.
- can find $\Phi$ when locally $X_t, Y_t$ “behave” like $R^*$
- Glauber dynamics converges locally to $R^*$
- Given $\Phi$ and convergence of Glauber dynamics we show rapid mixing
Key Results

- We don’t know a $\Phi$ that gives contraction for worst-case $X_t, Y_t$.
- can find $\Phi$ when locally $X_t, Y_t$ “behave” like $\omega^*$
- Glauber dynamics converges locally to $\omega^*$
- Given $\Phi$ and convergence of Glauber dynamics we show rapid mixing
Key Results

- We don’t know a $\Phi$ that gives contraction for worst-case $X_t, Y_t$.
- can find $\Phi$ when locally $X_t, Y_t$ “behave” like $\omega^*$
- Glauber dynamics converges locally to $\omega^*$
- Given $\Phi$ and convergence of Glauber dynamics we show rapid mixing
Theorem

Let $G$ be of girth $\geq 7$ and maximum degree $\Delta$, for $\Delta > \Delta_0$. Let $(X_t)$ be the Glauber dynamics with $\lambda < (1 - \delta)\lambda_c(\Delta)$.

For any vertex $v$, with probability $1 - \exp \left[ -\Delta / C \right]$, it holds that

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

where $t \geq Cn \log \Delta$. 
Key Results

- We don’t know a $\Phi$ that gives contraction for worst-case $X_t, Y_t$.
- Can find $\Phi$ when locally $X_t, Y_t$ “behave” like $\omega^*$
  - $\Phi$ is from the Jacobian of BP operator
- Glauber dynamics (approximately) converges locally to $\omega^*$
  - Locally Glauber dynamics behaves approximately like BP fixed points
- Given $\Phi$ and convergence of Glauber dynamics we show rapid mixing
There is a single disagreement at $v$
Rapid Mixing with uniformity
Dyer, Frieze, Hayes, Vigoda 2013

Run the chains for $Cn \log \Delta$ steps, “burn-in”
Rapid Mixing with uniformity
Dyer, Frieze, Hayes, Vigoda 2013

Run the chains for $Cn \log \Delta$ steps, “burn-in”
The disagreements spread in the graph during burn-in
Typically the disagreements do not escape the ball
Typically the disagreements do not escape the ball
Typically the ball has uniformity.
Interpolate and do path coupling for the pairs, 
... the pairs now “behave” like $\omega^*$
Interpolate and do path coupling for the pairs, 
\ldots the pairs now “behave” like \( \omega^* \) and \( \Phi \) gives convergence
$$\mathbb{E} \left[ \Phi(X_{C'n \log \Delta}, Y_{C'n \log \Delta}) \mid X_0, Y_0 \right] \leq (1 - \gamma) \Phi(X_0, Y_0)$$
Key Results

- We don’t know a $\Phi$ that gives contraction for worst-case $X_t, Y_t$.
- We can find $\Phi$ when $X, Y \sim \omega^*$
- Glauber dynamics converges locally to $\omega^*$
- Given $\Phi$ and convergence of Glauber dynamics we show rapid mixing
Key Results

- We don’t know a $\Phi$ that gives contraction for worst-case $X_t, Y_t$.
- We can find $\Phi$ when $X, Y \sim \omega^*$
- Glauber dynamics converges locally to $\omega^*$
- Given $\Phi$ and convergence of Glauber dynamics we show rapid mixing
Local uniformity I

\[ R(\sigma, v) = \prod_{w \sim v} \left( 1 - \frac{\lambda}{1 + \lambda} 1\{w \text{ unblocked by its children}\} \right), \]
Local uniformity I

$$R(\sigma, \nu) = \prod_{w \sim \nu} \left( 1 - \frac{\lambda}{1 + \lambda} \mathbf{1}_{\{w \text{ unblocked by its children}\}} \right),$$
\[ R(\sigma, v) = \prod_{w \sim v} \left( 1 - \frac{\lambda}{1 + \lambda} 1\{w \text{ unblocked by its children}\} \right), \]

\[ R(\sigma, v) = \Pr_{Y \sim \mu} [v \text{ is unblocked in } Y | v \notin Y, Y(S_2(v)) = \sigma(S_2(v))] \]
Local uniformity I

\[ R(\sigma, v) = \prod_{w \sim v} \left( 1 - \frac{\lambda}{1 + \lambda} 1\{w \text{ unblocked by its children}\} \right), \]

BP for Gibbs measure

Let \( G \) be of girth \( \geq 6 \) and maximum degree \( \Delta > \Delta_0 \). Let \( X \) be distributed as in \( \mu \) with \( \lambda < (1 - \delta)\lambda_c(\Delta) \).

Then for any vertex \( v \) with probability \( \geq 1 - \exp(-\Delta/C) \) it holds that

\[
\left| R(X, v) - \prod_{z \sim v} \left( 1 - \frac{\lambda}{1 + \lambda} R(X, z) \right) \right| < \gamma.
\]
Local uniformity I

\[ R(\sigma, v) = \prod_{w \sim v} \left( 1 - \frac{\lambda}{1 + \lambda} 1\{w \text{ unblocked by its children}\} \right), \]

BP for Glauber dynamics

Let \( G \) be of girth \( \geq 7 \) and maximum degree \( \Delta > \Delta_0 \). Let \( (X_t) \) be the Glauber dynamics with \( \lambda < (1 - \delta)\lambda_c(\Delta) \).

Then for any vertex \( v \) and any \( t > Cn \log \Delta \) with probability \( \geq 1 - \exp(-\Delta/C) \) it holds that

\[
\left| R(X_t, v) - \prod_{z \sim v} \left( 1 - \frac{\lambda}{1 + \lambda} \mathbb{E}_{tz}[R(X_{tz}, z)] \right) \right| < \gamma.
\]
Lemma

Let $G$ be of girth $\geq 7$ and maximum degree $\Delta > \Delta_0$. Let $(X_t)$ be the Glauber dynamics with $\lambda < (1 - \delta) \lambda_c(\Delta)$. For all $I = [t_0, t_1]$, where $t_0 = C_n \log \Delta$, for every $v \in V$ with probability $1 - (1 + |I|/n) \exp(-\Delta/C)$, we have that $(\forall t \in I)$ $|R(X_t, v) - \omega^*(v)| \leq \epsilon$. 

Hayes 2012
Lemma

Let $G$ be of girth $\geq 7$ and maximum degree $\Delta > \Delta_0$. Let $(X_t)$ be the Glauber dynamics with $\lambda < (1 - \delta)\lambda_c(\Delta)$.

For all $I = [t_0, t_1]$, where $t_0 = Cn \log \Delta$, for every $v \in V$ with probability $1 - (1 + |I|/n) \exp(-\Delta/C)$, we have that

$$(\forall t \in I) \quad |R(X_t, v) - \omega^*(v)| \leq \epsilon.$$
Lemma

Let $G$ be of girth $\geq 7$ and maximum degree $\Delta > \Delta_0$. Let $(X_t)$ be the Glauber dynamics with $\lambda < (1 - \delta)\lambda_c(\Delta)$. For all $I = [t_0, t_1]$, where $t_0 = Cn \log \Delta$, for every $v \in V$ with probability $1 - (1 + |I|/n) \exp(-\Delta/C)$, we have that

$$(\forall t \in I) \quad |R(X_t, v) - \omega^*(v)| \leq \epsilon.$$
Iterations in space and time
Convergence with $\Psi$

Potential function

$$\Psi(x) = (\lambda)^{-1} \arcsinh(\sqrt{\lambda x})$$
Convergence with $\Psi$

Potential function

$$\Psi(x) = (\lambda)^{-1} \text{arcsinh}(\sqrt{\lambda x})$$
Convergence with \( \Psi \)

Potential function

\[
\Psi(x) = (\lambda)^{-1} \text{arcsinh}(\sqrt{\lambda x})
\]
Iterations in space and time

Convergence with $\Psi$

Potential function

$$\Psi(x) = (\lambda)^{-1}\arcsinh(\sqrt{\lambda x})$$

Provided

- $t \in I'$ approximate BP equation hold in $B(v, R)$
- $\forall t \in I_{i+1}, u \in B(v, i + 1)$

$$|\Psi(R(X_t, u)) - \Psi(\omega^*(u))| \leq \alpha_{i+1}$$

$R \in R(v, u)$
Iterations in space and time

Convergence with $\Psi$

Potential function

$$\Psi(x) = (\lambda)^{-1} \text{arcsinh}(\sqrt{\lambda} x)$$

Provided

- $t \in I'$ approximate BP equation hold in $B(v, R)$
- $\forall t \in I_{i+1}, u \in B(v, i + 1)$

$$|\Psi(R(X_t, u)) - \Psi(\omega^*(u))| \leq \alpha_{i+1}$$

$\forall t \in I, u \in R(v, i)$

$$|\Psi(R(X_t, u)) - \Psi(\omega^*(u))| \leq (1 - 4i)\alpha_i$$
Iterations in space and time

Convergence with $\Psi$

Potential function

$$\Psi(x) = (\lambda)^{-1} \text{arcsinh}(\sqrt{\lambda} x)$$

Provided

- $t \in I'$ approximate BP equation hold in $B(v, R)$
- $\forall t \in I_{i+1}, u \in B(v, i + 1)$

$$|\Psi(R(X_t, u)) - \Psi(\omega^*(u))| \leq \alpha_{i+1}$$

$\forall t \in I_{i+1}, u \in B(v, i + 1)$

$$|\Psi(R(X_t, u)) - \Psi(\omega^*(u))| \leq (1 - \delta)\alpha_{i+1}$$
Convergence with $\Psi$

Potential function

$$\Psi(x) = (\lambda)^{-1} \text{arcsinh}(\sqrt{\lambda x})$$

Provided

- $t \in \mathcal{I}'$ approximate BP equation hold in $B(v, R)$
- $\forall t \in \mathcal{I}_{i+1}, u \in B(v, i + 1)$

$$|\Psi(R(X_t, u)) - \Psi(\omega^*(u))| \leq \alpha_{i+1}$$

$\forall t \in \mathcal{I}, u \in R(v)$

$$|\Psi(R(X_t, u)) - \Psi(\omega^*(u))| \leq (1 - \delta)\alpha_i$$
Iterations in space and time

Convergence with $\Psi$

Potential function

$$\Psi(x) = (\lambda)^{-1} \text{arcsinh}(\sqrt{\lambda x})$$

Provided

- $t \in I'$ approximate BP equation hold in $B(v, R)$
- $\forall t \in I_{i+1}, u \in B(v, i + 1)$

$$|\Psi(R(X_t, u)) - \Psi(\omega^*(u))| \leq \alpha_{i+1}$$

$\forall t \in I, u \in R(\nu_0)$

$$|\Psi(R(X_t, u)) - \Psi(\omega^*(u))| \leq 1 - \delta_i$$
Iterations in space and time

Convergence with $\Psi$

Potential function

$$\Psi(x) = (\lambda)^{-1} \text{arcsinh}(\sqrt{\lambda x})$$

Provided

- $t \in \mathcal{I}'$ approximate BP equation hold in $B(\nu, R)$
- $\forall t \in \mathcal{I}_{i+1}, u \in B(\nu, i + 1)$
  $$|\Psi(R(X_t, u)) - \Psi(\omega^*(u))| \leq \alpha_{i+1}$$
- $\forall t \in \mathcal{I}_i, u \in B(\nu, i)$
  $$|\Psi(R(X_t, u)) - \Psi(\omega^*(u))| \leq (1 - \delta)\alpha_{i+1}$$
Iterations in space and time

Convergence with $\Psi$

Potential function

$$\Psi(x) = (\lambda)^{-1} \arcsinh(\sqrt{\lambda x})$$

Provided

- $t \in I'$: approximate BP equation hold in $B(v, R)$
- $\forall t \in I_{i+1}, u \in B(v, i + 1)$
  
  $$|\Psi(R(X_t, u)) - \Psi(\omega^*(u))| \leq \alpha_{i+1}$$

- $\forall t \in I_i, u \in B(v, i)$
  
  $$|\Psi(R(X_t, u)) - \Psi(\omega^*(u))| \leq (1 - \delta)\alpha_{i+1}$$
Concluding Remarks

Rapid mixing for Glauber Dynamics

\[ \Delta > \Delta_0 \text{ and } \lambda \geq 7 \]

Approach by establishing uniformity proposing "Hamming weights"

Establish a novel connection between Path Coupling and Loopy BP.

This is important for both uniformity and Hamming weights.

Use experience from Glauber dynamics to analyze Loopy BP for graphs of girth \( \geq 6 \) in the uniqueness region.

The connection between Glauber dynamics and Loopy BP is deep.

Allows to establish uniformity and weights in a systematic way.
Concluding Remarks

- Rapid mixing for Glauber Dynamics
  - $G$ max degree $\Delta > \Delta_0$ and girth $\geq 7$
  - $\lambda$ in uniqueness
Concluding Remarks

- Rapid mixing for Glauber Dynamics
  - $\Delta > \Delta_0$ and girth $\geq 7$
  - $\lambda$ in uniqueness

- Approach
Concluding Remarks

- **Rapid mixing for Glauber Dynamics**
  - $G$ max degree $\Delta > \Delta_0$ and girth $\geq 7$
  - $\lambda$ in uniqueness

- **Approach**
  - by establishing *uniformity*
Concluding Remarks

- **Rapid mixing for Glauber Dynamics**
  - \( G \) max degree \( \Delta > \Delta_0 \) and girth \( \geq 7 \)
  - \( \lambda \) in uniqueness

- **Approach**
  - by establishing *uniformity*
  - proposing “Hamming weights”
Concluding Remarks

- Rapid mixing for Glauber Dynamics
  - $G$ max degree $\Delta > \Delta_0$ and girth $\geq 7$
  - $\lambda$ in uniqueness

- Approach
  - by establishing *uniformity*
  - proposing “Hamming weights”

- Establish a novel connection between Path Coupling and Loopy BP
  - this is important for both uniformity and Hamming weights
Concluding Remarks

- Rapid mixing for Glauber Dynamics
  - $G$ max degree $\Delta > \Delta_0$ and girth $\geq 7$
  - $\lambda$ in uniqueness

- Approach
  - by establishing uniformity
  - proposing “Hamming weights”

- Establish a novel connection between Path Coupling and Loopy BP
  - this is important for both uniformity and Hamming weights

- Use experience from Glauber dynamics to analyze Loopy BP
  - for graphs of girth $\geq 6$ in the uniqueness region
Concluding Remarks

- Rapid mixing for Glauber Dynamics
  - $G$ max degree $\Delta > \Delta_0$ and girth $\geq 7$
  - $\lambda$ in uniqueness

- Approach
  - by establishing *uniformity*
  - proposing “Hamming weights”

- Establish a novel connection between Path Coupling and Loopy BP
  - this is important for both uniformity and Hamming weights

- Use experience from Glauber dynamics to analyze Loopy BP
  - for graphs of girth $\geq 6$ in the uniqueness region

- The connection between Glauber dynamics and Loopy BP is deep
Concluding Remarks

- Rapid mixing for Glauber Dynamics
  - $G$ max degree $\Delta > \Delta_0$ and girth $\geq 7$
  - $\lambda$ in uniqueness
- Approach
  - by establishing *uniformity*
  - proposing “Hamming weights”
- Establish a novel connection between Path Coupling and Loopy BP
  - this is important for both uniformity and Hamming weights
- Use experience from Glauber dynamics to analyze Loopy BP
  - for graphs of girth $\geq 6$ in the uniqueness region
- The connection between Glauber dynamics and Loopy BP is deep
  - Allows to establish uniformity and weights in a *systematic way*
THANK YOU!