Lifted Probabilistic Inference for Asymmetric Graphical Models

Guy Van den Broeck and Mathias Niepert

Jan 28, 2015, AAAI

Take-Away Message

Two problems:

1. Lifted inference gives **exponential speedups** in **symmetric** graphical models.
   But what about real-world **asymmetric** problems?

2. When there are **many variables**, MCMC is **slow**.
   How to sample quickly in large graphical models?

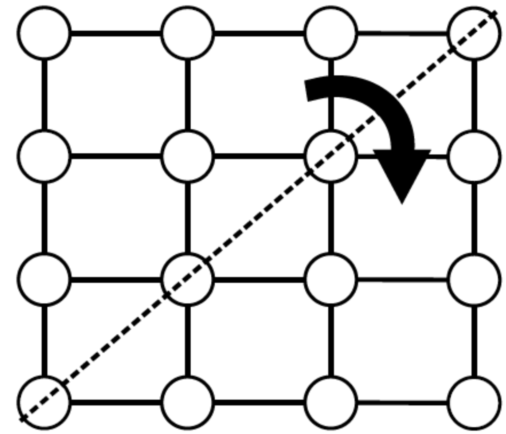
One solution: Exploit **approximate symmetries**!

Approximate Symmetries

- Symmetry g: $\Pr(\mathbf{x}) = \Pr(\mathbf{x}^g)$

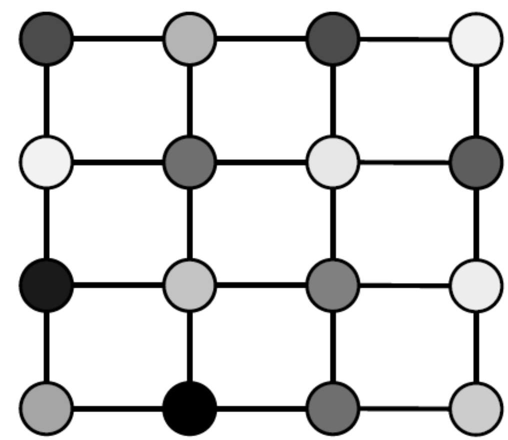
  E.g. Ising model without external field

$$\Pr\begin{pmatrix} 0&0&1&1 \\ 1&0&1&0 \\ 1&1&1&1 \\ 0&1&0&0 \end{pmatrix} = \Pr\begin{pmatrix} 0&1&1&0 \\ 1&1&0&0 \\ 0&1&1&1 \\ 0&1&0&1 \end{pmatrix}$$



- Approximate symmetry g: $\Pr(\mathbf{x}) \approx \Pr(\mathbf{x}^g)$

  E.g. Ising model with external field



Orbital Metropolis Chain: Algorithm

- Given symmetry group G (approx. symmetries)
- Orbit $\mathbf{x}^G$ contains all states approx. symm. to **x**
- In state **x**:

1. Select **y** uniformly at random from $\mathbf{x}^G$
2. Move from **x** to **y** with probability $\min\left(\frac{\Pr(\boldsymbol{y})}{\Pr(\boldsymbol{x})}, 1\right)$
3. Otherwise: stay in **x** (reject)
4. Repeat

Orbital Metropolis Chain: Analysis

- ✓ Pr(.) is stationary distribution
- ✓ Many variables change (fast mixing)
- ✓ Few rejected samples:

$$\Pr(\boldsymbol{y}) \approx \Pr(\boldsymbol{x}) \Rightarrow \text{ min}\left(\frac{\Pr(\boldsymbol{y})}{\Pr(\boldsymbol{x})}, 1\right) \approx 1$$

Is this the perfect proposal distribution?

Orbital Metropolis Chain: Analysis

- ✓ Pr(.) is stationary distribution
- ✓ Many variables change (fast mixing)
- ✓ Few rejected samples:

$$\Pr(\boldsymbol{y}) \approx \Pr(\boldsymbol{x}) \Rightarrow \min\left(\frac{\Pr(\boldsymbol{y})}{\Pr(\boldsymbol{x})}, 1\right) \approx 1$$

Is this the perfect proposal distribution?

- ✗ Not irreducible…
  Can never reach 0100 from 1101.

Lifted Metropolis-Hastings: Algorithm

- Given an **orbital Metropolis chain $\mathbf{M_S}$** for Pr(.)
- Given a **base Markov chain $\mathbf{M_B}$** that
  - is irreducible and aperiodic
  - has stationary distribution Pr(.)

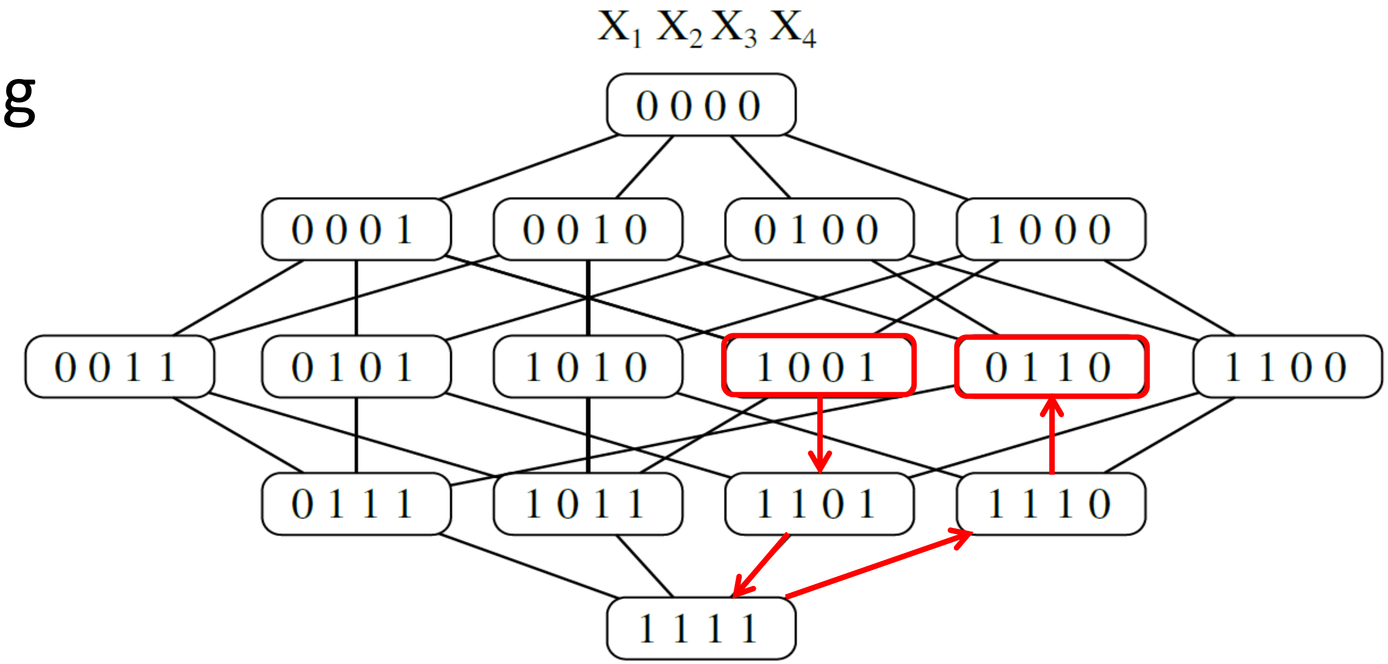
  (e.g., Gibbs chain or MC-SAT chain)
- In state **x**:
  1. With probability α, apply the kernel of M_B
  2. Otherwise apply the kernel of M_S

Lifted Metropolis-Hastings: Analysis

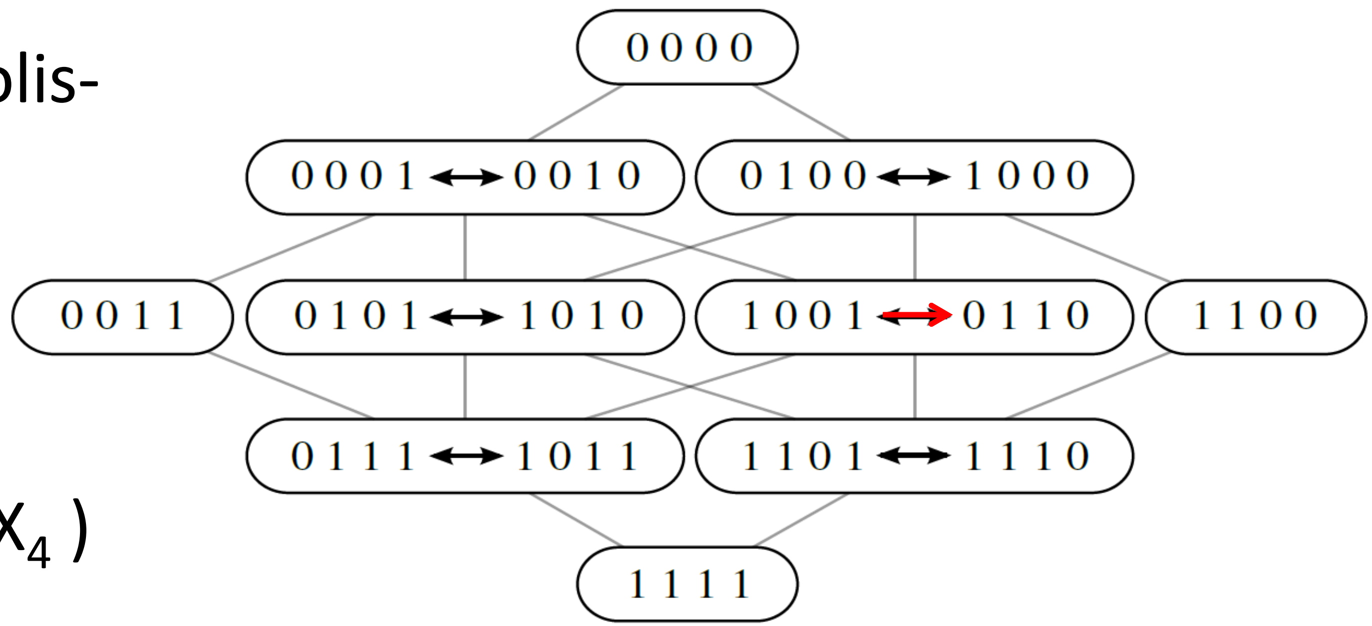
Theorem [Tierney 1994]:
A mixture of Markov chains is irreducible and aperiodic if at least one of the chains is irreducible and aperiodic .

- ✓ Pr(.) is stationary distribution
- ✓ Many variables change (fast mixing)
- ✓ Few rejected samples
- ✓ Irreducible
- ✓ Aperiodic

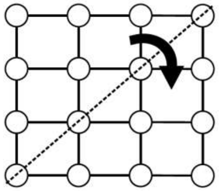
Gibbs Sampling



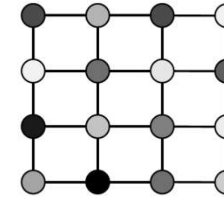
Lifted Metropolis-Hastings

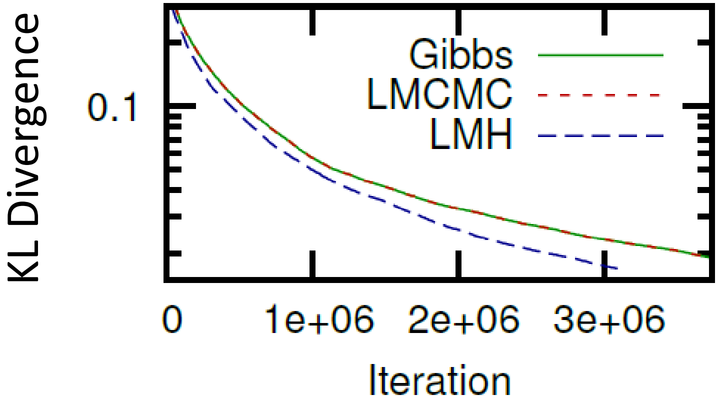


$G = (X_1 X_2)(X_3 X_4)$

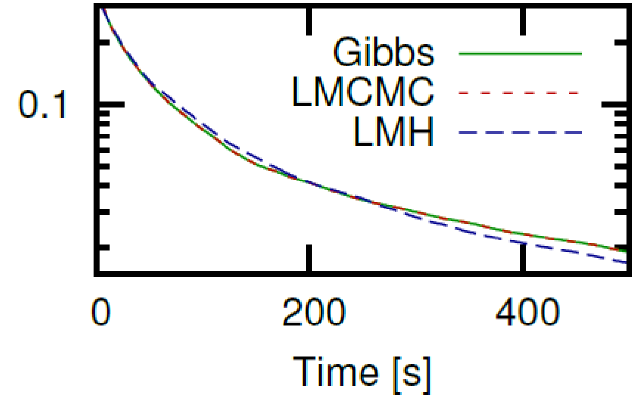


Example: Grid Models

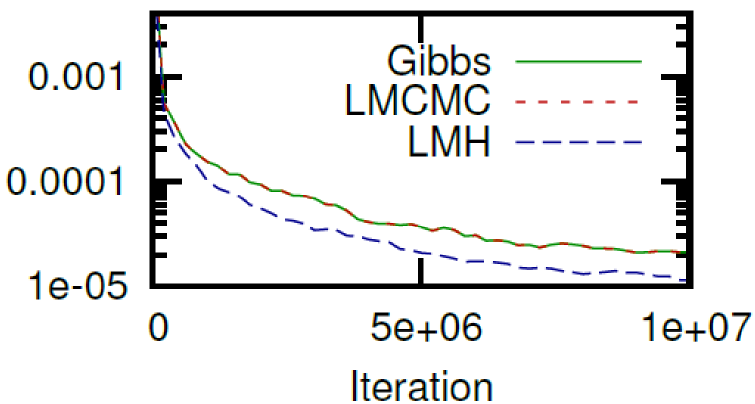




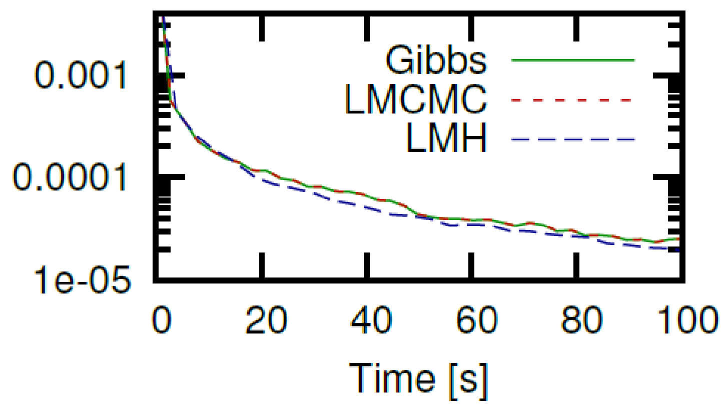
(a) Ising - Iterations



(b) Ising - Time



(c) Chimera - Iterations



(d) Chimera - Time

Example: Statistical Relational Model

- WebKB: Classify pages given links and words
- Very large Markov logic network

$$1.3 \quad \texttt{Page}(\mathsf{x}, \mathsf{Faculty}) \Rightarrow \texttt{HasWord}(\mathsf{x}, \mathsf{Hours})$$

$$1.5 \quad \texttt{Page}(\mathsf{x}, \mathsf{Faculty}) \wedge \texttt{Link}(\mathsf{x}, \mathsf{y}) \Rightarrow \texttt{Page}(\mathsf{y}, \mathsf{Course})$$

and 5000 more …

- No symmetries with evidence on Link or Word
- Where do approx. symmetries come from?

Over-Symmetric Approximations

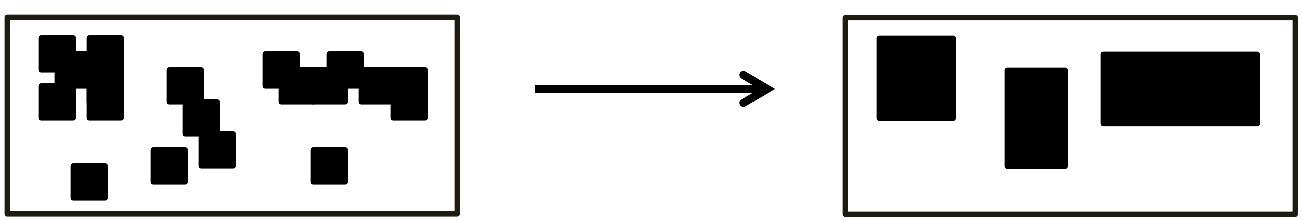
- OSA makes model more symmetric
- E.g., low-rank Boolean matrix factorization

Link (“aaai.org”, “google.com”)
Link (“google.com”, “aaai.org”)
Link (“google.com”, “gmail.com”)
Link (“ibm.com”, “aaai.org”)

→

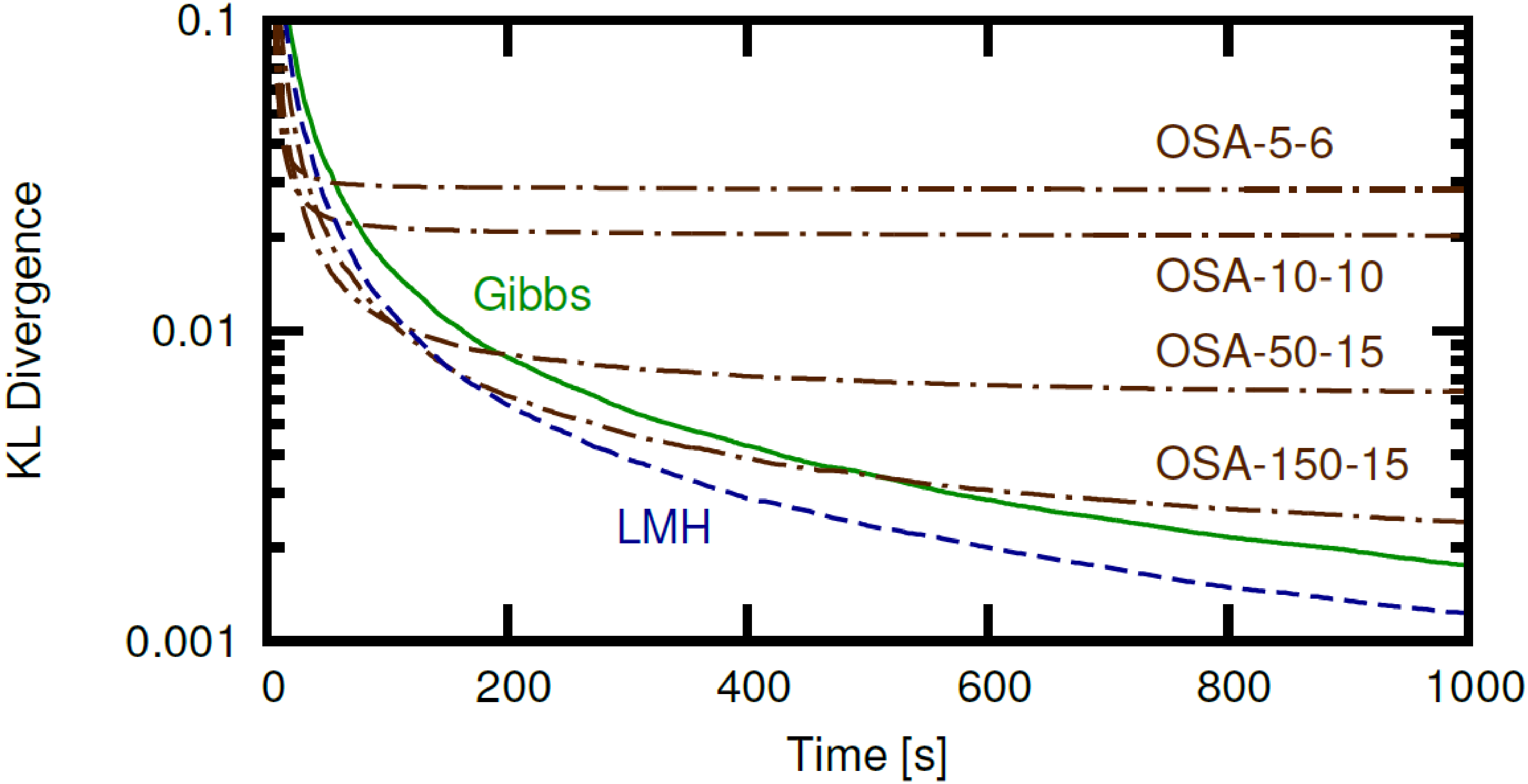
Link (“aaai.org”, “google.com”)
Link (“google.com”, “aaai.org”)
- ~~Link (“google.com”, “gmail.com”)~~
\+ Link (“aaai.org”, “ibm.com”)
Link (“ibm.com”, “aaai.org”)

google.com and ibm.com become symmetric!

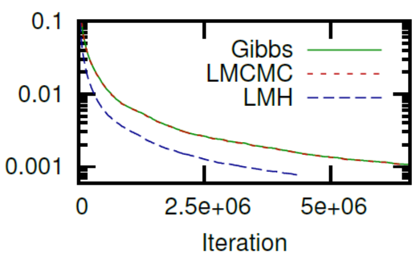


[Van den Broeck & Darwiche ‘13], [Venugopal and Gogate ‘14], [Singla, Nath and Domingos ‘14]

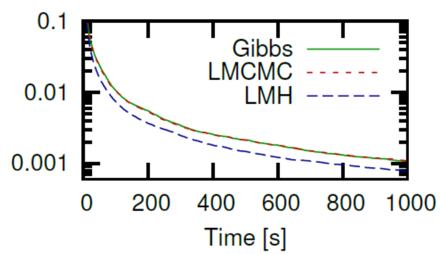
Experiments: WebKB



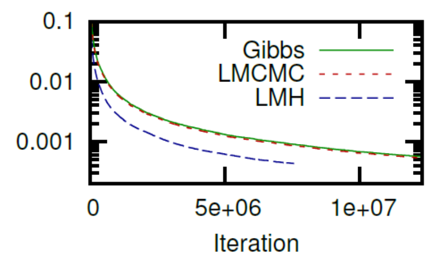
Experiments: WebKB



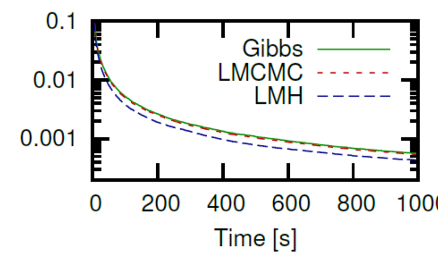
(a) Texas - Iterations



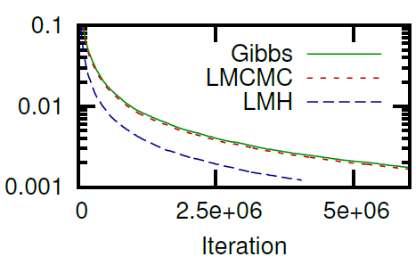
(b) Texas - Time



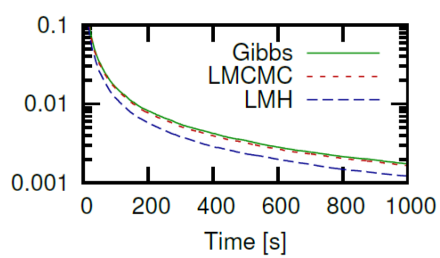
(a) Cornell - Iterations



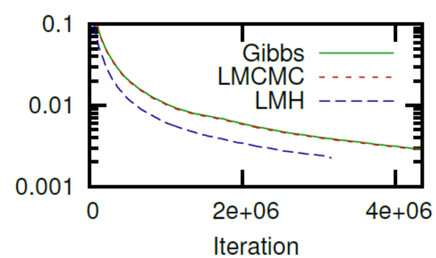
(b) Cornell - Time



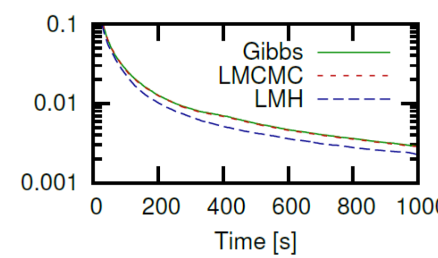
(c) Washington - Iterations



(d) Washington - Time



(c) Wisconsin - Iterations



(d) Wisconsin - Time

Conclusions

- Lifted Metropolis Hastings
  - works on any graphical model
  - exploits approximate symmetries
  - does not require any exact symmetries
  - converges to the true marginals
  - mixes faster (changes many variables per iteration)
  - has low rejection rate
- Practical lifted inference algorithm
- Need more research on over-symmetric approximations!

Thank you