Sampling from Probabilistic Submodular Models

Alkis Gotovos Hamed Hassani Andreas Krause

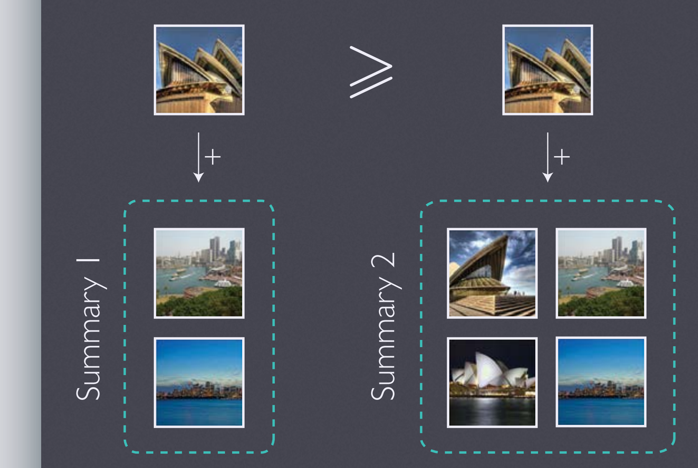
ETH zürich

Motivation

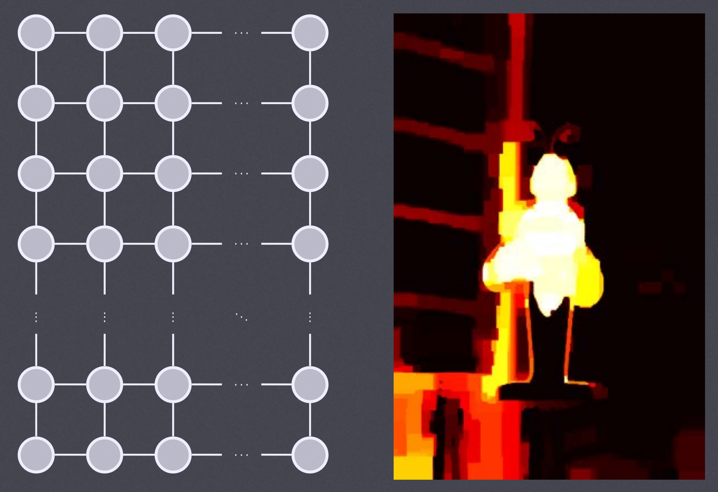
Image collection summarization

Foreground / background segmentation

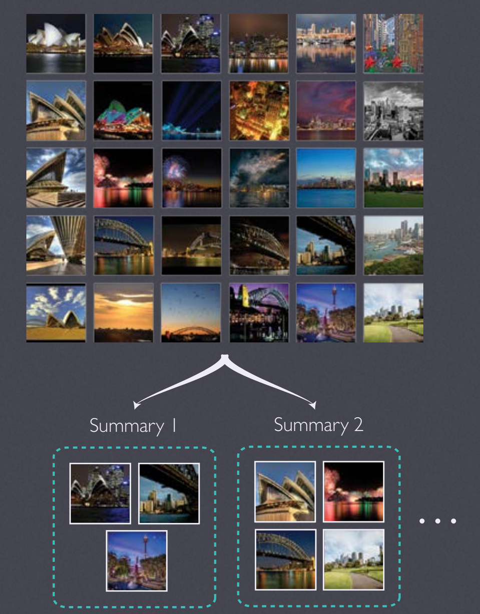
Submodularity



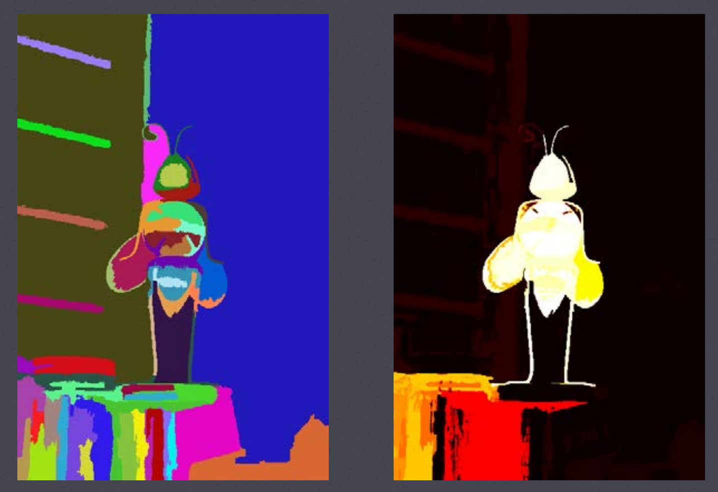
Pairwise models



Sampling summaries



Higher-order models



[Djolonga and Krause, '15]

Use submodular functions in probabilistic models

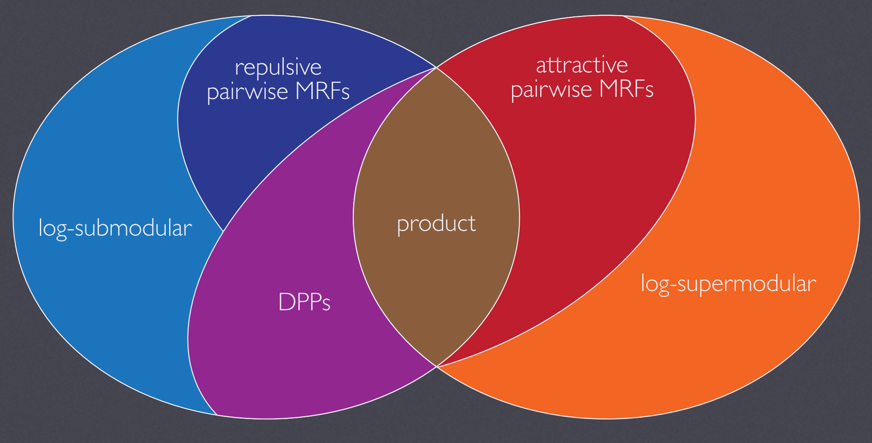
Equip existing models with higher-order interactions

Probabilistic Submodular Models

Probabilistic Submodular Models

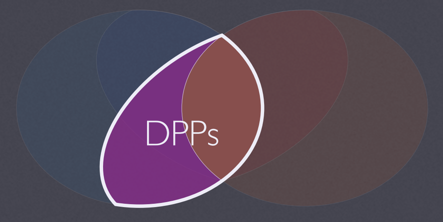
$$p(S) = \frac{1}{Z}\exp(F(S))$$

PSMs	MRFs
◦ Ground set V with $\lvert V \rvert = n$	◦ Binary random vector $X = (X_1, \dots, X_n)$
◦ Sub- or supermodular function $F : 2^V \to \mathbb{R}$	◦ Set of factors $\phi_i : \{0,1\}^{\mathcal{C}_i} \to \mathbb{R}$
◦ Distribution over subsets $p(S) \propto \exp(F(S))$	◦ Distribution over binary vectors $p(X) \propto \exp\left(\sum_i \phi_i(X_{\mathcal{C}_i})\right)$



Inference

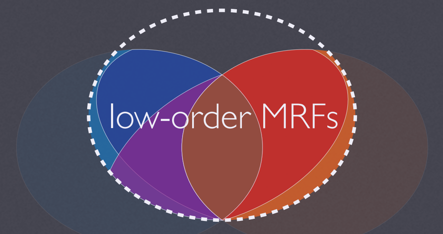
EXACT



◦ Tractable only for limited subclasses

◦ #P-hard even for Ising models

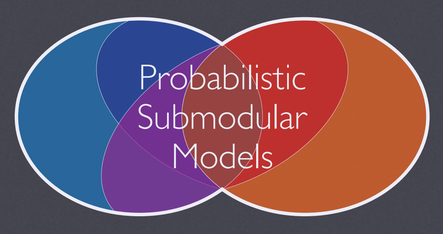
BP, MF, ...



◦ Extensively studied model class

◦ Complexity exponential in model order

L-FIELD



◦ Variational approach for general PSMs [Djolonga and Krause, '14]

What about sampling?

Gibbs Sampling

Start at S_0

For $t = 1, 2, \dots$

◦ Select random $v \in V$

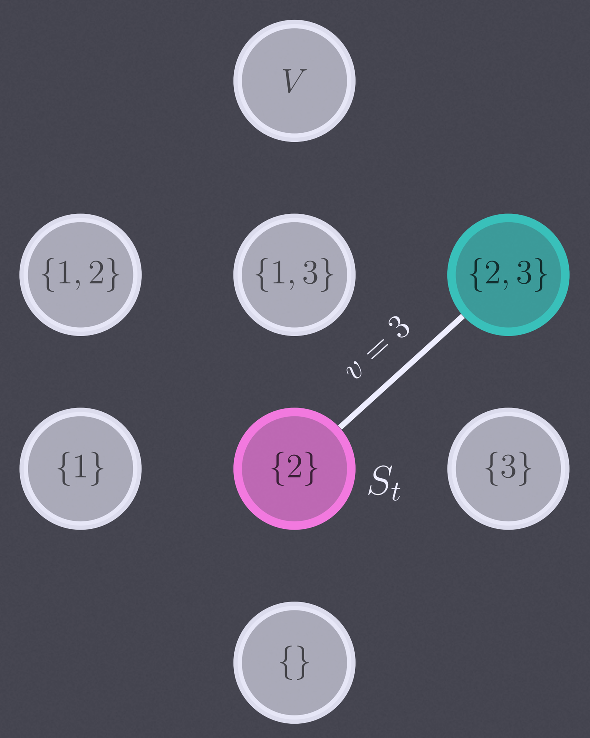
◦ $\Delta \leftarrow F(S_t \cup \{v\}) - F(S_t \setminus \{v\})$

◦ $p_{\text{add}} \leftarrow e^{\Delta} / (1 + e^{\Delta})$

◦ Flip biased coin

p_{add} | $1 - p_{\text{add}}$

$S_{t+1} \leftarrow S_t \cup \{v\}$ $S_{t+1} \leftarrow S_t \setminus \{v\}$



Does the Markov chain converge?

How long does it take to converge?

Total variation distance

$$d(t) = \max\{d_{\text{tv}}(\mathbb{P}_{S_t}, \pi) \mid S_0 \in \Omega\}$$

Mixing time

$$t_{\text{mix}}(\epsilon) = \min\{t \mid d(t) \le \epsilon\}$$

Under mild assumptions (ergodicity),

$$d(t) \xrightarrow{t \to \infty} 0$$

We establish sufficient conditions for sub-exponential mixing of the Gibbs sampler on PSMs

Polynomial-time Mixing

"Distance" from modularity

$$\zeta_F := \max_{A,B \subseteq V} \left|F(A) + F(B) - F(A \cup B) - F(A \cap B)\right|$$

submodular		constant		modular		normalized monotone submodular
$F(S)$	$=$	c	$+$	$\sum_{v \in S} m_v$	$+$	$f(S)$

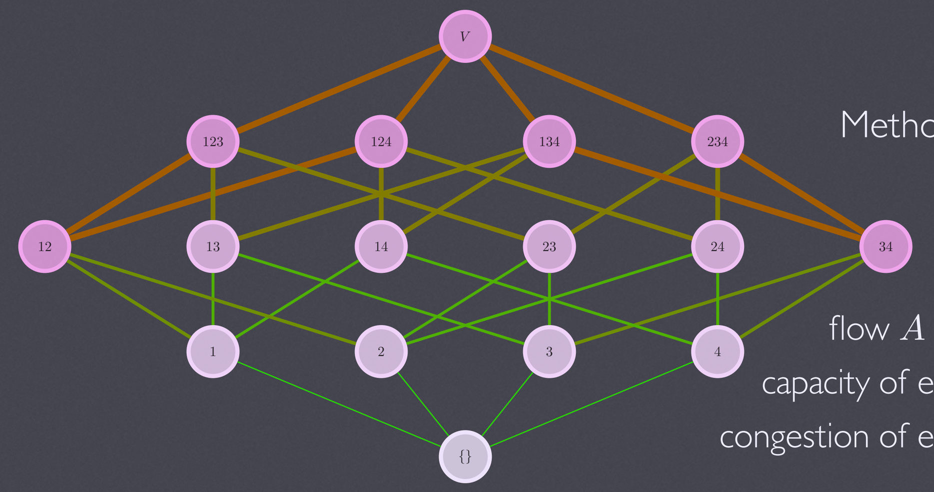
Theorem 1

For any set function F, the mixing time of the Gibbs sampler is bounded by

$$t_{\text{mix}}(\epsilon) \le 2n^2 \exp(2\zeta_F) \log(\epsilon p_{\min})^{-1}.$$

For any submodular or supermodular set function F, the mixing time of the Gibbs sampler is bounded by

$$t_{\text{mix}}(\epsilon) \le 2n^2 \exp(\zeta_f) \log(\epsilon p_{\min})^{-1}.$$



Method of canonical paths [Sinclair, '92]

flow $A \to B$	—	$p(A)p(B)$
capacity of edge e	—	transition probability
congestion of edge e	—	total flow / capacity

Fast Mixing

Influence of $r \in V$ on $v \in V$

$$\gamma_F(v; r \mid S) := \left|\Delta_F(v \mid S) - \Delta_F(v \mid S \cup \{r\})\right|$$

Maximum total influence

$$\gamma_F := \max_{\substack{r \in V \\ S \subseteq V}} \sum_{v \in V} \tanh\left(\frac{1}{2}\gamma_F(v; r \mid S)\right)$$

Theorem 2

For any submodular or supermodular set function F, if $\gamma_f < 1$, the mixing time of the Gibbs sampler is bounded by

$$t_{\text{mix}}(\epsilon) \le \frac{1}{1 - \gamma_f} n \left(\log n + \log \epsilon^{-1}\right).$$

We call f decomposable if $f(S) = \sum_{i \in [L]} f_i(S)$

Lemma 1

For any decomposable submodular function f,

$$\gamma_f \le \frac{1}{2}\theta_f \lambda_f.$$

$$\theta_f := \max_{v \in V} \sum_{i \in [L]} \sqrt{f_i(v)}$$

$$\lambda_f := \max_{i \in [L]} \sum_{v \in V} \sqrt{f_i(v)}$$

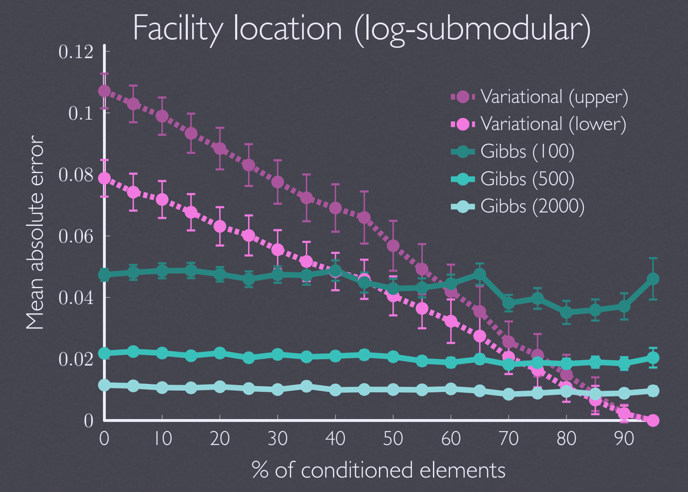
Evaluation

◦ Compare against variational inference [Djolonga and Krause, '14]

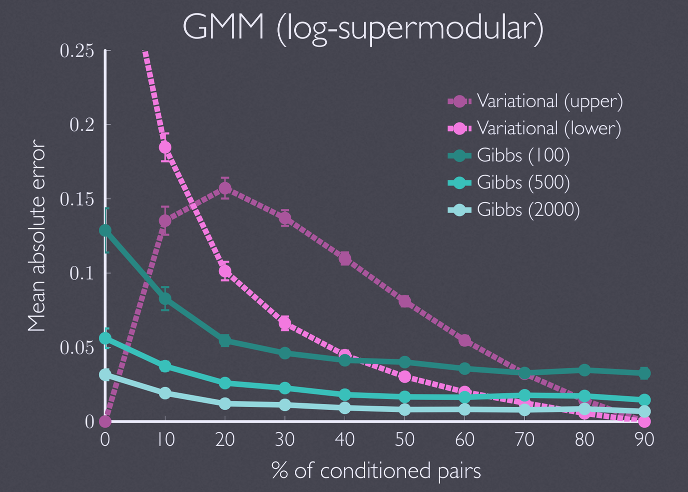
◦ Compute $p(v \mid S)$

◦ $\lvert V \rvert = 20 \longrightarrow$ exact marginals

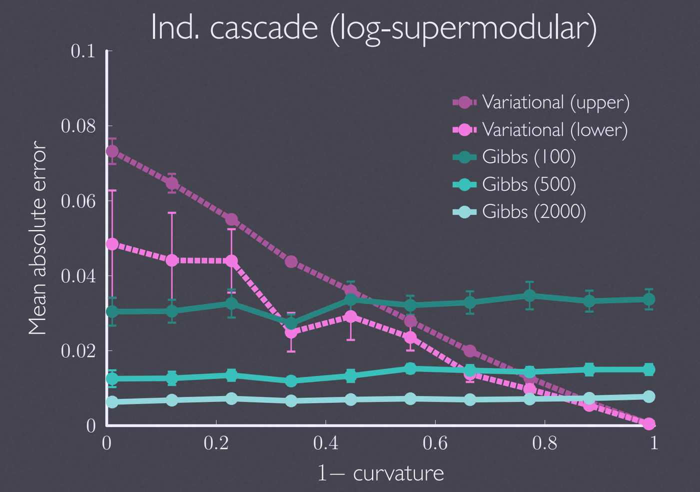
Facility location (log-submodular)



GMM (log-supermodular)



Ind. cascade (log-supermodular)



[1] Josip Djolonga and Andreas Krause. From MAP to marginals: Variational inference in bayesian submodular models. In Neural Information Processing Systems, 2014.

[2] Josip Djolonga and Andreas Krause. Scalable variational inference in log-supermodular models. In International Conference on Machine Learning, 2015.

[3] Martin Dyer, Leslie Ann Goldberg, and Mark Jerrum. Matrix norms and rapid mixing for spin systems. Annals of Applied Probability, 2009.

[4] Patrick Rebeschini and Amin Karbasi. Fast mixing for discrete point processes. In Conference on Learning Theory, 2015.

[5] Alistair Sinclair. Improved bounds for mixing rates of markov chains and multicommodity flow. Combinatorics, Probability and Computing, 1992.