Clamping Variables and Approximate Inference  
Adrian Weller and Tony Jebara

Summary
We address the problem of *marginal inference* for undirected graphical models - estimating the partition function $Z$ and marginal probability distributions. We focus on binary pairwise (Ising) models, e.g. vision, RBMs, or social networks. Combining clamping of variables with *approximate variational inference* we obtain:
- Strong theoretical results (middle column)
- Promising empirical results (right column)

Background: What is clamping?
$Z$ can be split into two parts: clamping variable $X_i$ to each of $\{0, 1\}$, then add the two sub-partition functions:
$$Z = Z|_{X_i = 0} + Z|_{X_i = 1}$$
After clamping a variable, remove it from the graph:
- If remaining sub-models are acyclic then can find sub-partition functions efficiently (Bethe approximation is exact on trees)
- If not, (a) can repeat until acyclic, or (b) settle for approximate inference on sub-models

Clamping
$$Z_B := Z|_{X_i \rightarrow \theta} = Z|_{X_i = 0} + Z|_{X_i = 1}$$
But sub-partition functions are exact, hence $\text{LHS} = Z$

Let $\nu(G)$ be minimum size of a *feedback vertex set* (set of vertices such that deleting them renders graph acyclic; $\nu \geq \text{treewidth} - 1$)

**Theorem (result is tight)**
$$Z \leq 2^\nu Z_B$$

Attractive models: *Master Theorem* (strongest result)
An attractive model has all variables positively coupled:
- For any variable $X_i$ and $x \in \{0, 1\}$, let $q_i = q(x_i = 1)$ and
  $$\log Z_B(x) = \max_{q \in Q} \left[ \theta \cdot q + S_B(q) \right]$$
- $Z_B(x)$ is ‘Bethe partition function constrained to singleton $q_i = x$’
- Define new function,
  $$A(q_i) := \log Z_B(q_i \cdot S_B(q))$$
- By considering derivatives of the Bethe free energy, and how the optimum constrained to singleton $q_i$ varies with $q_i$, we show
  **Theorem (strongest result for attractive models)**
  For an attractive binary pairwise model, $A(q)$ is convex

Attractive models: Consequences of Master Theorem
Lower bound on $Z_B^{(l)}$ leads to lower bound on $Z$

**Theorem** (clamp and sum can only increase Bethe)

For an attractive binary pairwise model and any $X_i$, $Z_B \leq Z_B^{(l)}$
Then with similar proof to result above for general models, **Corollary** (lower bound on $Z$, first proved by Ruozzi, 2012)
For an attractive binary pairwise model, $Z_B \leq Z$

$$Z_B \leq Z_B^{(l)} \leq Z_B^{(l)} \leq \cdots \leq Z$$

Experiments
We investigate error of $Z_B^{(l)} = Z_B|_{X_i = 0} + Z_B|_{X_i = 1}$ compared to error of original $Z_B$, various methods of choosing variable to clamp $X_i$:
- *best Clamp* best improvement in error of $Z$ in hindsight
- *worst Clamp* worst improvement in error of $Z$ in hindsight
- *avg Clamp* average performance
- *maxW* max sum of incident edge weights $\sum_j |W_{ij}|$
- *Mpower* more sophisticated, tries to break heavy cycles

Potentials drawn at random: unary $\theta_i \sim U[-2, 2]$, edge $W_{ij} \sim U[0, W_{\text{max}}]$ for general, $W_{ij} \sim U[0, W_{\text{max}}]$ for attractive models

**Attractive random graph $n = 10$, $p = 0.5$**

**General random graph $n = 10$, $p = 0.5$**

As $n$ grows, still helpful even just to clamp one variable (see paper)