This talk is about

- Belief propagation (BP)
  - Heuristic for inference in graphical models
  - Empirically very successful, but
    - Theoretical underpinning of its strength/limitation is still elusive
    - And primary motivation for this work

- Rest of the talk (using one example)
  - What is BP?
  - Complexity of BP
  - Correctness of BP
  - Connection to a conjecture of Alon-Tarsi ‘85
Example: counting independent set

- Given an undirected graph $G=(V,E)$
  - $V$ being vertices: $n$ of them
  - $E$ being edge-set: assume max vertex degree $d$ (constant)
- Independent set of $G$
  - Subset of $V$ so that no two nodes in it are connected as per $E$

- Goal:
  - Count the number of independent sets of $G$

\[n=10\]
\[d=5\]
Example: counting independent set

- Counting independent sets of $G$ is equivalent to
  - Computing marginals in a Markov Random Field (MRF)

- Let $X = [X_v]$ be distribution over $\{0,1\}^n$ such that
  \[
P(X = x) \propto \prod_{(u,v) \in E} (1 - x_u x_v) \quad \text{where} \quad x \in \{0,1\}^n
  \]

- Then, the normalization constant of the distribution
  - Equals the number of independent sets of $G$
  - Which is (computationally) equivalent to computing
    - $P(X_v = 1)$ for all $v$ : BP is a heuristic for computing this.
Belief propagation (BP)

- Belief propagation is
  - Parallel implementation of dynamic programing for a tree graph
  - Applied to any graph
Belief propagation (BP)

- Belief propagation is
  - Tree approx. of parallel dynamic programing

\[
\frac{P(x_w=1)}{P(x_w=0)} = \frac{Z(x_w=1)}{Z(x_w=0)} = \frac{Z(x_v=0)}{Z(x_v=0) + Z(x_v=1)} = \frac{1}{1 + \prod_{u \neq v} Z_u(x_v=1)} \frac{1}{Z_u(x_v=0)}
\]

\[Z_u(x_v=1)\] 
\[
\text{# of ind. set with } x_v=1 \text{ in this sub-tree}
\]
Belief propagation (BP)

- Belief propagation is
  - Tree approx. of parallel dynamic programming

\[ m_{v \rightarrow w} = \frac{1}{1 + \prod_{u \in \mathcal{N}(v)} m_{u \rightarrow v}} \]

\[ m_{u \rightarrow v} = \frac{P_v(x_v = 1)}{P_v(x_v = 0)} \]

\[ m_{v \rightarrow w} = \frac{P_v(x_w = 1)}{P_v(x_w = 0)} \]
Belief propagation (BP)

- Belief propagation is
  - Tree approx. of parallel dynamic programing

\[
m_{v \rightarrow w}^{t+1} = \frac{1}{1 + \prod_{u \in N(v)} m_{u \rightarrow v}^{t}}
\]

\[
p_t(x_v = 1) = \frac{\prod_{u \in N(v)} m_{u \rightarrow v}^{t}}{1 + \prod_{u \in N(v)} m_{u \rightarrow v}^{t}}
\]

\[
m_{v \rightarrow w}^{0} = \frac{1}{2}
\]
Belief propagation (BP)

- Belief propagation is tree approx. of dyn. prgm.
  - Exact for tree graph, but
  - What about arbitrary graph?

- BP is an iterative algorithm
  - Q1. Does it have fixed points?
  - Q2. If so, what are they?
  - Q3. Does algorithm converge to them?
  - Q4. And, how good are they?
Q1: existence of fixed point

- Recall, BP iteration

\[ m^0_{v \rightarrow w} = \frac{1}{2} \quad \text{and} \quad m^{t+1}_{v \rightarrow w} = \frac{1}{1 + \prod_{u \in N(v) \setminus w} m^t_{u \rightarrow v}} \quad \text{for all} \quad t \geq 0 \quad (v, w) \in E \]

- It can be checked that

\[ m^t_{v \rightarrow w} \in \left[ \frac{1}{2}, 1 \right] \]

- Therefore, by Brouwer’s fixed point theorem

\( \text{Fixed point(s) does exist} \)

(Yedidia-Freeman-Weiss ’01; also Wainwright-Jaakkola-Willsky ‘03)
Q2: characterization of fixed point

- Recall Tree graph

\[
\log \left( \frac{\text{# of ind}}{\text{set}} \right) = H(x_1, \ldots, x_n) = H(x_1) + \sum_{v=2}^{n} H(x_v|x_1)
\]

for any tree

\[
= \sum_{v \in \mathcal{V}} H(x_v) - \sum_{(u,v) \in \mathcal{E}} I(x_u; x_v)
\]

\[
= -\sum_{v \in \mathcal{V}} (d_v-1) H(x_v) + \sum_{(u,v) \in \mathcal{E}} H(x_u, x_v)
\]
Q2: characterization of fixed point

- For any tree graph $G=(V, E)$

$$\log \left( \frac{\text{ind}}{\text{sets}} \right) = - \sum_{v \in V} (d_v - 1) \mathcal{H}(x_v) + \sum_{(u, v) \in E} \mathcal{H}(x_u, x_v)$$
Q2: characterization of fixed point

- For any tree graph $G = (V, E)$

\[
\log \left( \frac{\text{# ind sets}}{\text{sets}} \right) = - \sum_{v \in V} \left( d_v - 1 \right) H(x_v) + \sum_{(u,v) \in E} H(x_u, x_v)
\]

Let $x_v = \mathbb{P}(x_v = 1)$.

By independent set constraint $\mathbb{P}(x_u = 1, x_v = 1) = 0$ for all $(u,v) \in E$.

Therefore

\[
\mathbb{P}(x_u = 1, x_v = 0) = x_u, \quad \mathbb{P}(x_u = 0, x_v = 1) = x_v
\]

\[
\mathbb{P}(x_u = x_v = 0) = 1 - x_u - x_v
\]
Q2: characterization of fixed point

- For any tree graph $G=(V, E)$
  $$\log\left(\frac{\text{# ind. sets}}{\text{sets}}\right) = -\sum_{v \in V} (d_v - 1) H(x_v) + \sum_{(u, v) \in E} H(x_u, x_v)$$

Let $x_v = P(x_v = 1)$.

By independent set constraint $P(x_u = 1, x_v = 1) = 0$ for all $(u, v) \in E$.

Therefore
  $$P(x_u = 1, x_v = 0) = x_u, P(x_u = 0, x_v = 1) = x_v$$
  $$P(x_u = x_v = 0) = 1 - x_u - x_v$$

Hence:
  $$\log\left(\frac{\text{# ind. sets}}{\text{sets}}\right) = F\left((x_v)_{v \in V}\right) = -\sum_{v \in V} x_v \ln x_v - \sum_{(u, v) \in E} (1 - x_u - x_v) \ln (1 - x_u - x_v) + \sum_{v \in V} (d_v - 1)(1 - x_v) \ln (1 - x_v)$$
Q2: characterization of fixed point

- For any tree graph \( G=(V,E) \)
  - Let \( x = [x_v] \) be fixed point marginals
  - Then the log (\# ind. sets) is \( F(x) \) where

\[
F(x) = - \sum_{v \in V} x_v \ln x_v - \sum_{(v,w) \in E} (x_v - x_v) \ln (1-x_v) - \sum_{v \in V} (d_v-1) x_v \ln (1-x_v)
\]

- This is called the Bethe Approximation
- The fixed point \( x = [x_v] \) solves (for tree)

\[
\text{maximize} \quad F(y) \\
\text{over} \quad y \in [0,1]^h \\
\text{subject to} \quad y_u + y_v \leq 1 \quad \text{for all} \quad (u,v) \in E
\]

That is, \( \nabla F(x) = 0 \).
Q2: characterization of fixed point

- For any graph $G=(V,E)$
  - $x = [x_v]$ corresponds to (non-trivial) BP fixed point marginals
  - If and only if $\nabla F(x) = 0$

(Yedidia, Freeman and Weiss ‘01)

- Thus, if BP converges then
  - Then it solves certain fixed point equation

- Next question is that of convergence
Q3: Convergence of BP

- For any graph with max degree 5
  - The so called “correlation decay” implies convergence
    (Kelly ‘87, Tatikonda-Jordan ’00)

- Recently, for graphs with degree larger than 5
  - Hardness of counting independent set has been established
    (Sly ‘10)

- Question remains
  - Convergence of BP beyond 5 graphs w. deg 5 not known
    - Is BP also trying to solve hard problem beyond degree 5?
    - If not, can we find convergent (poly-time) BP-like algorithm?
Finding BP fixed points

- Given graph $G=(V,E)$
  - Find $x = [x_v]$ so that

$$\nabla F(x) = 0$$

over $x \in [0,1]^n$ for all $x_u + x_v \leq 1$ for all $(u,v) \in E$

- Main result: using a BP-like algorithm
  - For any graph $G=(V,E)$, BP fixed point can be found approximately in number of iterations that scales
    - Quadratic in $n$ (number of nodes)
    - Exponentially in $d$ (degree)
    - Inverse square dependence on the approx. accuracy
Finding BP fixed points

More precisely, for any graph $G = (V, E)$

$\hat{x} \in [0,1]^n$ can be found so that

$\|\nabla F(\hat{x})\| \leq \varepsilon$

with number of iterations $O(n^2d^84^d\varepsilon^{-2}(\log(1/\varepsilon)))$

for any $\varepsilon \in (0,1)$.

Prior work by Yuille ’02

- Provides convergent algorithm
- But does not establish any rate of convergence
- We overcome this challenge by interesting use of gradient-like algo.
Finding BP fixed points

- Algorithm is essentially the standard “gradient” algorithm
  - With careful choice of step-size, initial condn & answer
  - Non-triviality is in it’s analysis

Choose \( \alpha(t) = \sqrt{\frac{2^{d+1}}{(d^2 + 6d + 2) 2^t}} \)

Initially \( y(0) = [y_v(0)] = [\frac{1}{4}] \)

For \( 1 \leq t \leq T = \Theta\left( n^d d^8 4^d \frac{1}{\epsilon^2} \log \left( \frac{1}{\epsilon} \right) \right) \)

\[ y(t+1) = y(t) + \alpha(t) \nabla F(y(t)) \]

Produce \( \hat{x} \) as \( \epsilon \)-approximate fixed point by picking \( \hat{x} = y(t) \) with smallest \( \| \nabla F(y(t)) \| \)
Q4: Correctness of Bethe approximation

**Two issues**

- Robustness of Bethe approximation
  - That is, understanding the effect of approximate fixed point
- How good is the Bethe approximation?
Q4: Correctness of Bethe approximation

- **Two issues**
  - Robustness of Bethe approximation
    - That is, understanding the effect of approximate fixed point
  - How good is the Bethe approximation?

- **Robustness**
  - Chertkov and Chernyak ‘06 showed that
    \[
    \frac{Z}{\exp(F(x))} = \mathcal{G}(G, x)
    \]
    depends on cycles of \(G\) and the BP fixed pt \(x\).
Q4: Correctness of Bethe approximation

- Two issues
  - Robustness of Bethe approximation
    - That is, understanding the effect of approximate fixed point
  - How good is the Bethe approximation?

- Robustness
  - Chertkov and Chernyak ‘06 showed that
    \[
    \frac{Z}{\exp(F(x))} = \mathcal{G}(G, x)
    \]
  - We show that
    \[
    \frac{Z}{\exp(F(x))} = \mathcal{G}(G, \hat{x})(1 \pm \epsilon)^n
    \]
Q4: Correctness of Bethe approximation

- Two issues
  - Robustness of Bethe approximation
    - That is, understanding the effect of approximate fixed point
  - How good is the Bethe approximation?

- Correctness of Bethe approximation
  - Known for graph with max degree 5
  - And large enough girth
    (Bandyopadhyay-Gamarnik ‘06, Weitz ‘06)
Q4: Correctness of Bethe approximation

- Two issues
  - Robustness of Bethe approximation
    - That is, understanding the effect of approximate fixed point
  - How good is the Bethe approximation?

- Correctness of Bethe approximation
  - We show that
    - For any graph with max degree $d$
    - If girth is larger than $8d \log n$ then
    - Approximate BP fixed points can find
      - number of independent set approximately
    - That is, it is a PTAS
A dilemma?

- Alon and Tarsi ’85 made a combinatorial conjecture
  - About covering (bridgeless) graphs with cycles

- Assuming this conjecture
  - Our results imply that for a random 3 regular graph
    \[ Z = (1.545...)^n + C(\varepsilon) \] with prob. \(1-\varepsilon\)
  - This is too sharp as generically we expect
    - error term to scale with \(n\) (not independent of \(n\))

- So either AT conjecture is false or
  - Bethe approximation is too good!
    (or 3 regular graph is v. well behaved)
Discussion

- This work has been about
  - Complexity and correctness of belief propagation
  - That is, the Bethe approximation

- The Bethe approximation (or BP fixed point)
  - Is easy to find for sparse graphs
  - It’s complexity, however, depends exponentially on degree

- The Bethe approximation
  - Is robust
  - Correct beyond degree 5 graphs; but at cost of larger girth

- Interesting connection to Alon-Tarsi’s conjecture