Message Passing for Collective Graphical Models

Tao Sun\textsuperscript{1}, Daniel Sheldon\textsuperscript{1,2}, Akshat Kumar\textsuperscript{3}

\textsuperscript{1}University of Massachusetts Amherst
\textsuperscript{2}Mount Holyoke College
\textsuperscript{3}Singapore Management University

ICML 2015, Lille
Motivation: Aggregate Data

Individual data is hard to collect because
- unable to identify individuals
- privacy concerns
- and etc.

but aggregate data is readily available . . .
Motivation: Aggregate Data

Individual data is hard to collect because

- unable to identify individuals $\implies$ insect counts
- privacy concerns $\implies$ census data
- and etc.

but aggregate data is readily available . . .
Motivation: Aggregate Data

Individual data is hard to collect because

- unable to identify individuals $\implies$ insect counts
- privacy concerns $\implies$ census data
- and etc.

but aggregate data is readily available . . .

Collective Graphical Models: tool for inference and learning when only (noisy) aggregate data is available.
**Example: Bird Migration**

- **eBird project**: bird watchers submit checklists containing bird counts and environmental covariates.
Example: Bird Migration

- **eBird project**: bird watchers submit checklists containing bird counts and environmental covariates.
- **Goal**: use eBird data to fit models of bird migration.
Example: Bird Migration

- **eBird project**: bird watchers submit checklists containing bird counts and environmental covariates.
- **Goal**: use eBird data to fit models of bird migration.
- **Problem**: eBird shows evidence of migration, but doesn’t directly model migration.
Example: Bird Migration

- **eBird project**: bird watchers submit checklists containing bird counts and environmental covariates.
- **Goal**: use eBird data to fit models of bird migration.
- **Problem**: eBird shows evidence of migration, but doesn’t directly model migration.
Want a dynamic/transition model for bird migration.
Individual bird: Markov chain

\[ X_t \in \{1, \ldots, L\}, \text{location at time } t \]

\[ \theta \]

\[ X_1 \rightarrow X_2 \rightarrow \cdots \rightarrow X_T \]
Generative Model for eBird Data

Individual bird: Markov chain

\[ X_1 \rightarrow X_2 \rightarrow \cdots \rightarrow X_T \]

- \( X_t \in \{1, \cdots, L\} \), location at time \( t \)
- Parameter \( \theta \)
Generative Model for eBird Data

\( M \) i.i.d. individuals

\[
X_1^m \xrightarrow{} X_2^m \xrightarrow{} \ldots \xrightarrow{} X_T^m
\]

\( m = 1, \ldots, M \)
Generative Model for eBird Data

*M i.i.d. individuals*

\[ X_1^m \rightarrow X_2^m \rightarrow \cdots \rightarrow X_T^m \]

\[ m = 1, \ldots, M \]

- \( X_t^m \in \{1, \cdots, L\} \), location of bird \( m \) at time \( t \)
Aggregate counts

\[ X^m_1 \rightarrow X^m_2 \rightarrow \ldots \rightarrow X^m_T \]

\[ m = 1, \ldots, M \]

\[ n_1 \rightarrow A \begin{array}{c} 54 \\ 32 \end{array} \rightarrow n_2 \rightarrow n_T \]
Generative Model for eBird Data

Aggregate counts

\[ X^m_1 \rightarrow X^m_2 \rightarrow \ldots \rightarrow X^m_T \]

\[ m = 1, \ldots, M \]

\[ n_t(x_t) = \# \text{ birds in location } x_t \text{ at time } t \]

\[ n_1, A = 54, B = 32 \]

\[ n_2 \]

\[ n_T \]
eBird: noisy aggregate counts

\[
X^m_1 \rightarrow X^m_2 \rightarrow \ldots \rightarrow X^m_T \\
\text{where } m = 1, \ldots, M
\]

\[
\begin{array}{c|c|c}
\text{n}_1 & A & 54 \\
 & B & 32 \\
\hline
\text{y}_1 & A & 10 \\
 & B & 6 \\
\end{array}
\]

\[
\begin{array}{c|c|c}
\text{n}_2 & & \\
\hline
\text{y}_2 & & \\
\end{array}
\]

\[
\begin{array}{c|c|c}
\text{n}_T & & \\
\hline
\text{y}_T & & \\
\end{array}
\]

\[p(y|n)\] noise model, e.g., \[y_t(x_t)|n_t(x_t) \sim \text{Poisson}(\alpha_{n_t}(x_t))\]
eBird: noisy aggregate counts

\[ X_1^m \rightarrow X_2^m \rightarrow \ldots \rightarrow X_T^m \quad m = 1, \ldots, M \]

- \( y_1 \): A 10, B 6
- \( n_1 \): A 54, B 32
- \( y_T \): A 10, B 6
- \( n_T \): A 54, B 32

- \( p(y \mid n) \) noise model
Generative Model for eBird Data

eBird: noisy aggregate counts

\[ X_1^m \rightarrow X_2^m \rightarrow \ldots \rightarrow X_T^m \]

\[ m = 1, \ldots, M \]

\[ n_1 \quad \begin{array}{c|c} A & 54 \\ B & 32 \end{array} \]

\[ n_2 \]

\[ y_1 \quad \begin{array}{c|c} A & 10 \\ B & 6 \end{array} \]

\[ y_2 \]

\[ y_T \]

- \( p(y \mid n) \) noise model
- e.g., \( y_t(x_t) \mid n_t(x_t) \sim \text{Poisson}(\alpha n_t(x_t)) \)
Problem Statement

\[
X_1^m \rightarrow X_2^m \rightarrow \ldots \rightarrow X_T^m \\
\quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad m = 1, \ldots, M
\]

- posterior inference over hidden variables
- learn parameter \( \theta \)
Generative Model for Collective Human Mobility

- Transition counts
- Corrupt data by adding noise to maintain differential privacy
Transition counts
Corrupt data by adding noise to maintain differential privacy
General CGMs

- Any discrete (undirected) graphical model inside the plate
- For each individual $p(x; \theta) = \frac{1}{Z(\theta)} \prod_{i \sim j} \psi_{ij}(x_i, x_j; \theta)$
- (Noisy) observations of aggregate counts (contingency table)
Background: *Collective Graphical Models*

Analytically marginalize away individuals  

[Sheldon and Dietterich NIPS 2011]
Background: *Collective Graphical Models*

**Analytically marginalize away individuals**  
[Sheldon and Dietterich NIPS 2011]
Background: *Collective Graphical Models*

- $n_{t,t+1}(x_t, x_{t+1}) = \# \text{ birds that fly from location } x_t \text{ to } x_{t+1} \text{ at time } t$
- $n_{t,t+1}$ are sufficient statistics for parameter $\theta$
- Closed form probability model $p(n)$ for trees / junction trees
  
  [Sundberg 1975, Liu et al. ICML 2014]
Formal Problem Statement

\[ n_1, n_2, n_3, \ldots, n_{T-1}, n_T \]

\[ p(n) \]

Marginal inference:
\[ p(n | y) \]

MAP inference:
\[ \max_n p(n | y) \]

Learning:
\[ \max_{\theta} p(y | \theta) \]

Exact inference is intractable even for trees

[Sheldon et al. ICML 2013]

≈⇒ Approximate inference
Formal Problem Statement

- **Marginal inference:** $p(n \mid y)$
- **MAP inference:** $\max_n p(n \mid y)$
- **Learning:** $\max_{\theta} p(y \mid \theta)$
Formal Problem Statement

- Marginal inference: $p(n \mid y)$
- MAP inference: $\max_n p(n \mid y)$
- Learning: $\max_\theta p(y \mid \theta)$

Exact inference is intractable even for trees  
[Sheldon et al. ICML 2013]
Formal Problem Statement

Marginal inference: $p(n \mid y)$

MAP inference: $\max_n p(n \mid y)$

Learning: $\max_\theta p(y \mid \theta)$

Exact inference is intractable even for trees  [Sheldon et al. ICML 2013]

⇒ Approximate inference!
Our Contributions

- Highlight a connection between the approximate MAP inference in CGMs and the marginal inference in standard graphical models.
Our Contributions

- Highlight a connection between the **approximate MAP inference** in *CGMs* and the **marginal inference** in *standard graphical models*
- Develop a novel message passing algorithm
Our Contributions

- Highlight a connection between the approximate MAP inference in CGMs and the marginal inference in standard graphical models.
- Develop a novel message passing algorithm.
- It’s faster and more accurate than all previous approaches.
Our Contributions

- Highlight a connection between the approximate MAP inference in CGMs and the marginal inference in standard graphical models.
- Develop a novel message passing algorithm.
- It’s faster and more accurate than all previous approaches.
- Evaluate on a synthetic bird migration benchmark and a novel human mobility application.
Connection between Two Inference Problems

Approximate MAP inference for CGMs
Connection between Two Inference Problems

Marginal inference for standard GMs \(\Rightarrow\) Approximate MAP inference for CGMs
Connection between Two Inference Problems

Marginal inference for standard GMs \iff Approximate MAP inference for CGMs

\[ \downarrow \] [Yedidia et al. NIPS 2001]

Belief Propagation

???
Connection between Two Inference Problems

Marginal inference for standard GMs

\[ \approx \]

Approximate MAP inference for CGMs

\[ \approx \]

Belief Propagation

[Yedidia et al. NIPS 2001]

Nonlinear Belief Propagation (NLBP)
Marginal inference for standard GMs

Approximate MAP inference for CGMs
Marginal inference for standard GMs

\[
\min_{z \in L_1} F_B(z) \quad \text{Bethe Free Energy}
\]

Approximate MAP inference for CGMs

\[
\min_{z \in L_M} F_{\text{CGM}}(z)
\]
### Marginal inference for standard GMs

\[
\min_{z \in \mathbb{L}_1} F_B(z) \quad \text{Bethe Free Energy}
\]

### Approximate MAP inference for CGMs

\[
\min_{z \in \mathbb{L}_M} F_{\text{CGM}}(z) \quad (n \rightarrow z)
\]
Marginal inference for standard GMs

\[
\min_{\mathbf{z}} F_B(\mathbf{z}) = E_B(\mathbf{z}) - H_B(\mathbf{z}) \quad \text{Bethe Free Energy}
\]

Approximate MAP inference for CGMs

\[
\min_{\mathbf{z}} F_{\text{CGM}}(\mathbf{z}) = E_{\text{CGM}}(\mathbf{z}) - H_B(\mathbf{z})
\]
Marginal inference for standard GMs

\[
\min_{\mathbf{z} \in \mathbb{L}_1} F_B(\mathbf{z}) = E_B(\mathbf{z}) - H_B(\mathbf{z}) \quad \text{Bethe Free Energy}
\]

Approximate MAP inference for CGMs

\[
\min_{\mathbf{z} \in \mathbb{L}_M} F_{\text{CGM}}(\mathbf{z}) = E_{\text{CGM}}(\mathbf{z}) - H_B(\mathbf{z})
\]

Bethe entropy:

\[
H_B(\mathbf{z}) = - \sum_{(i,j) \in E} \sum_{x_i, x_j} z_{ij}(x_i, x_j) \log z_{ij}(x_i, x_j) + \sum_{i \in V} (\nu_i - 1) \sum_{x_i} z_i(x_i) \log z_i(x_i)
\]
Marginal inference for standard GMs

\[
\min_{z \in \mathbb{L}_1} F_B(z) = E_B(z) - H_B(z) \quad \text{Bethe Free Energy}
\]

\[
E_B(z) = - \sum_{(i,j) \in E} \sum_{x_i, x_j} z_{ij}(x_i, x_j) \log \psi_{ij}(x_i, x_j)
\]

Approximate MAP inference for CGMs

\[
\min_{z \in \mathbb{L}_M} F_{\text{CGM}}(z) = E_{\text{CGM}}(z) - H_B(z)
\]

Bethe entropy:

\[
H_B(z) = - \sum_{(i,j) \in E} \sum_{x_i, x_j} z_{ij}(x_i, x_j) \log z_{ij}(x_i, x_j) + \sum_{i \in V} (\nu_i - 1) \sum_{x_i} z_i(x_i) \log z_i(x_i)
\]
Marginal inference for standard GMs

\[
\min_{\mathbf{z} \in \mathbb{L}_1} F_B(\mathbf{z}) = E_B(\mathbf{z}) - H_B(\mathbf{z}) \quad \text{Bethe Free Energy}
\]

\[
E_B(\mathbf{z}) = - \sum_{(i,j) \in E} \sum_{x_i, x_j} z_{ij}(x_i, x_j) \log \psi_{ij}(x_i, x_j)
\]

Approximate MAP inference for CGMs

\[
\min_{\mathbf{z} \in \mathbb{L}_M} F_{CGM}(\mathbf{z}) = E_{CGM}(\mathbf{z}) - H_B(\mathbf{z})
\]

\[
E_{CGM}(\mathbf{z}) = - \sum_{(i,j) \in E} \sum_{x_i, x_j} z_{ij}(x_i, x_j) \log \psi_{ij}(x_i, x_j) - \log p(\mathbf{y} | \mathbf{z})
\]

Bethe entropy:

\[
H_B(\mathbf{z}) = - \sum_{(i,j) \in E} \sum_{x_i, x_j} z_{ij}(x_i, x_j) \log z_{ij}(x_i, x_j) + \sum_{i \in V} (\nu_i - 1) \sum_{x_i} z_i(x_i) \log z_i(x_i)
\]
Nonlinear Energy BP (NLBP) for CGMs

Repeat (1)–(3) in any order until convergence

1. Update edge potential based on current marginals $z$

$$\hat{\psi}_{ij}(x_i, x_j) = \exp \left\{ - \frac{\partial E(z)}{\partial z_{ij}(x_i, x_j)} \right\}$$

2. Standard message updates

$$m_{ij}(x_j) \propto \sum_{x_i} \hat{\psi}_{ij}(x_i, x_j) \prod_{k \in N(i) \setminus j} m_{ki}(x_i) \prod_{l \in N(j) \setminus i} m_{lj}(x_j)$$

3. Update marginals

$$z_{ij}(x_i, x_j) \propto \hat{\psi}_{ij}(x_i, x_j) \prod_{k \in N(i) \setminus j} m_{ki}(x_i) \prod_{l \in N(j) \setminus i} m_{lj}(x_j)$$

Identical to BP except potential updated in each iteration. Reduces to standard BP for linear energies.
Nonlinear Energy BP (NLBP) for CGMs

**Repeat (1)–(3) in any order until convergence**

1. **Update edge potential based on current marginals** $z$
   
   $$\hat{\psi}_{ij}(x_i, x_j) = \exp \left\{ - \frac{\partial E(z)}{\partial z_{ij}(x_i, x_j)} \right\}$$

2. **Standard message updates**

   $$m_{ij}(x_j) \propto \sum_{x_i} \hat{\psi}_{ij}(x_i, x_j) \prod_{k \in N(i) \setminus j} m_{ki}(x_i)$$
Nonlinear Energy BP (NLBP) for CGMs

**Repeat** (1)–(3) in any order until convergence

1. **Update edge potential based on current marginals** $\mathbf{z}$

   \[ \hat{\psi}_{ij}(x_i, x_j) = \exp \left\{ - \frac{\partial E(z)}{\partial z_{ij}(x_i, x_j)} \right\} \]

2. **Standard message updates**

   \[ m_{ij}(x_j) \propto \sum_{x_i} \hat{\psi}_{ij}(x_i, x_j) \prod_{k \in N(i) \setminus j} m_{ki}(x_i) \]

3. **Update marginals**

   \[ z_{ij}(x_i, x_j) \propto \hat{\psi}_{ij}(x_i, x_j) \prod_{k \in N(i) \setminus j} m_{ki}(x_i) \prod_{l \in N(j) \setminus i} m_{lj}(x_j) \]
Nonlinear Energy BP (NLBP) for CGMs

Repeat (1)–(3) in any order until convergence

1. Update edge potential based on current marginals $z$

$$\hat{\psi}_{ij}(x_i, x_j) = \exp \left\{ - \frac{\partial E(z)}{\partial z_{ij}(x_i, x_j)} \right\}$$

2. Standard message updates

$$m_{ij}(x_j) \propto \sum_{x_i} \hat{\psi}_{ij}(x_i, x_j) \prod_{k \in N(i) \setminus j} m_{ki}(x_i)$$

3. Update marginals

$$z_{ij}(x_i, x_j) \propto \hat{\psi}_{ij}(x_i, x_j) \prod_{k \in N(i) \setminus j} m_{ki}(x_i) \prod_{l \in N(j) \setminus i} m_{lj}(x_j)$$

- Identical to BP except potential updated in each iteration
Nonlinear Energy BP (NLBP) for CGMs

Repeat (1)–(3) in any order until convergence

1. Update edge potential based on current marginals $z$

$$\hat{\psi}_{ij}(x_i, x_j) = \exp \left\{ - \frac{\partial E(z)}{\partial z_{ij}(x_i, x_j)} \right\}$$

2. Standard message updates

$$m_{ij}(x_j) \propto \sum_{x_i} \hat{\psi}_{ij}(x_i, x_j) \prod_{k \in N(i) \setminus j} m_{ki}(x_i)$$

3. Update marginals

$$z_{ij}(x_i, x_j) \propto \hat{\psi}_{ij}(x_i, x_j) \prod_{k \in N(i) \setminus j} m_{ki}(x_i) \prod_{l \in N(j) \setminus i} m_{lj}(x_j)$$

- Identical to BP except potential updated in each iteration
- Reduces to standard BP for linear energies $E(z)$
Feasibility-Preserving NLBP

- New version of NLBP that alternates between updating edge potentials and call to standard BP as a subroutine
Feasibility-Preserving NLBP

- New version of NLBP that alternates between updating edge potentials and call to standard BP as a subroutine

- Advantages
Feasibility-Preserving NLBP

- New version of NLBP that alternates between updating edge potentials and call to standard BP as a subroutine

- Advantages
  - Very simple
Feasibility-Preserving NLBP

- New version of NLBP that alternates between updating edge potentials and call to standard BP as a subroutine

- Advantages
  - Very simple
  - Maintains feasibility (consistent marginals)
Feasibility-Preserving NLBP

- New version of NLBP that alternates between updating edge potentials and call to standard BP as a subroutine

- Advantages
  - Very simple
  - Maintains feasibility (consistent marginals)
  - Much faster in practice
**Theorem.** If NLBP converges, it finds a constrained stationary point of the CGM optimization problem. For tree-structured models with log-concave noise, it finds a global minimum.
Theorem. If NLBP converges, it finds a constrained stationary point of the CGM optimization problem. For tree-structured models with log-concave noise, it finds a global minimum.

- Open question whether NLBP converges even for trees
**Theorem.** If NLBP converges, it finds a constrained stationary point of the CGM optimization problem. For tree-structured models with log-concave noise, it finds a global minimum.

- Open question whether NLBP converges even for trees
- In practice sufficient damping always leads to convergence
Bird Migration Benchmark

[Sheldon et al. ICML 2013, Liu et al. ICML 2014]
Use CGM inference within E step of EM

L=6x6, w=[0.5,1,1,1]
Learning – Large Problem

$L=10 \times 10, w=[5, 10, 10, 10]$

- GENERIC
- NLBP−NAIVE
- NLBP−FEAS
- GCGM
Collective Human Mobility

6:00 AM

[Grid representation of human mobility]
Collective Human Mobility
Collective Human Mobility

6:00 AM

7:00 AM

8:00 AM
Collective Human Mobility

6:00 AM

7:00 AM

8:00 AM

9:00 AM

Tao Sun, Daniel Sheldon, Akshat Kumar
Collective Human Mobility

6:00 AM

7:00 AM

8:00 AM

9:00 AM

10:00 AM
Baseline: use noisy transition counts $y$ in place of $n$ to estimate $\theta$ (pretend there is no noise)

Our approach: infer $n$ with NLBP in E step of EM algorithm
Pairwise MAE – smaller is better!

![Graph showing performance with baseline and edge evidence.](image)

- **Performance**

- **Pairwise MAE** – smaller is better!
Want the inferred counts align with the true counts on the diagonal line.
Want the inferred counts align with the true counts on the diagonal line.
Want the inferred counts align with the true counts on the diagonal line.
Conclusion

- Highlight a connection between the *approximate MAP inference* in CGMs and *marginal inference* in standard graphical models.
- Develop a novel message passing algorithm for CGMs.
- It’s faster and more accurate than all previous approaches.
- Evaluate on a synthetic bird migration benchmark and a novel human mobility application.
Conclusion

- Highlight a connection between the *approximate MAP inference* in CGMs and *marginal inference* in standard graphical models.
- Develop a novel message passing algorithm for CGMs.
- It’s faster and more accurate than all previous approaches.
- Evaluate on a synthetic bird migration benchmark and a novel human mobility application.

Thanks!

Question?
Stable fixed points of loopy belief propagation are minima of the Bethe free energy.
*NIPS* 2003

Sundberg, R. (1975)
Some results about decomposable (or Markov-type) models for multidimensional contingency tables: distribution of marginals and partitioning of tests.

Yedidia, J. S., Freeman, W. T., Weiss, Y (2001)
Generalized belief propagation.
*NIPS* 2001

Sheldon, Daniel R and Dietterich, Thomas G (2011)
Collective graphical models.
*NIPS* 2011


\[ p(x_{1:10}) = \frac{1}{Z} \phi_1(x_1) \cdot \left( \prod_{t=1}^{9} \psi(x_t, x_{t+1}) \right) \cdot \phi_{10}(x_{10}) \]

\[ \psi(x_t, x_{t+1}) \propto \exp \left( - \frac{\|v_t - v_{t+1}\|^2}{(2\sigma^2)} \right) \]