# Graph Covers and Quadratic Minimization


Nicholas Ruozzi (Yale University)

Justin Thaler (Harvard University)

Sekhar Tatikonda (Yale University)

# Motivation

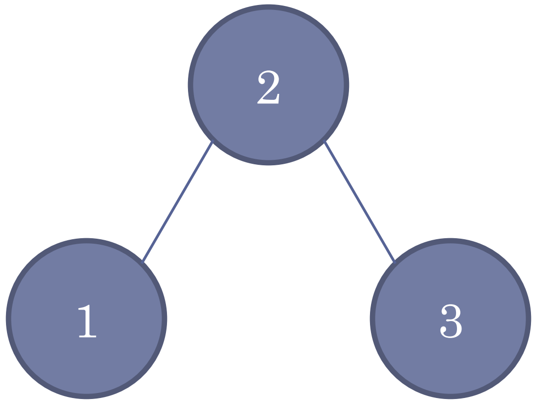
- No natural way to develop sufficient conditions for convergence of BP
- Many aspects of the algorithm are poorly understood for general problems:
  - Convergence
  - Correctness
  - Periodic behavior
- Quadratic minimization is easier
  - Closed form solution for the messages
  - Still exhibits many of the same phenomena

# Min-Sum

$$f(x_1, \ldots, x_n) = \sum_i \phi_i(x_i) \sum_\alpha \psi_\alpha(x_\alpha)$$

- The objective function factorizes as a sum of potentials
- $\alpha \subseteq \{1,\ldots,n\}$
- Pairwise if $|\alpha| = 2$

# Corresponding Graph




$$f(x_1, x_2, x_3) = \phi_1 + \phi_2 + \phi_3 + \psi_{23} + \psi_{12}$$

# MIN-SUM

- Messages at step n:

$$m_{i\to j}^{t}(x_j) = \min_{x_i} \phi_i(x_i) + \psi_{ij}(x_i, x_j) + \sum_{k\in\partial i - j} m_{k\to i}^{t-1}(x_i)$$

- Initial messages are chosen to be zero

## Computing Beliefs

- At each step, construct a set of beliefs:

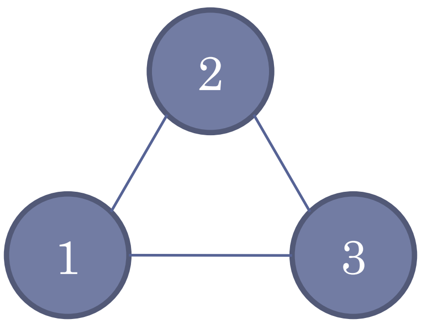
$$b_i^t(x_i) = \phi_i(x_i) + \sum_{j \in \partial i} m_{k \to i}^t(x_i)$$

- Estimate the optimal assignment as

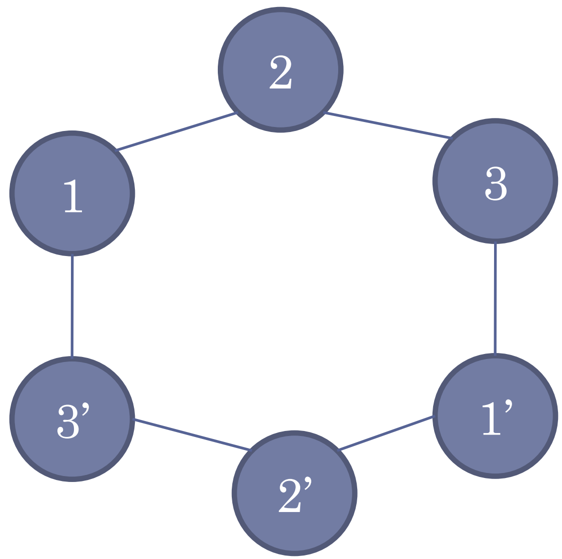
$$x_i^t = \arg\min_{x_i} b_i^t(x_i)$$

# Graph Covers

- A graph H covers a graph G if there is homomorphism from H to G that is a bijection on neighborhoods




Graph G

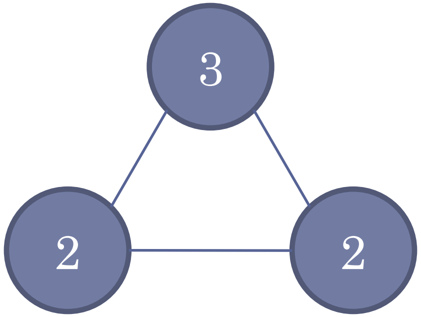



2-cover of G

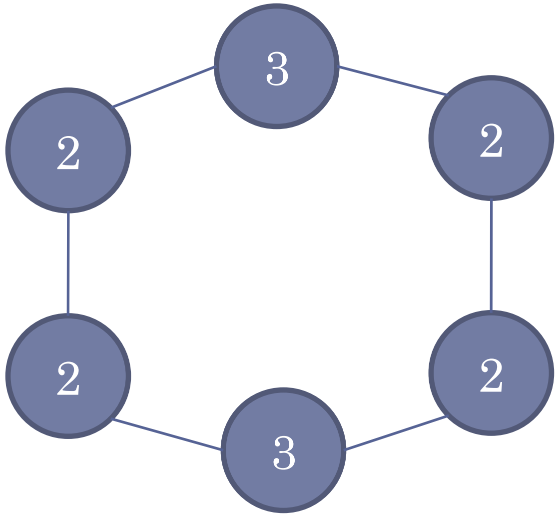
# Graph Covers

- Indistinguishability: for any cover and any choice of initial messages on the original graph, there exists a choice of initial messages on the cover such that the messages passed by min-sum are identical on both graphs
- Problems may arise if covers have different solutions than the original graph

# MWIS: A "Bad" Example




Graph G




2-cover of G

# Quadratic Minimization

$$f(x_1, ..., x_n) = \frac{1}{2}x^T\Gamma x - h^T x$$

- $\Gamma$ symmetric positive definite implies a unique minimum at $x = \Gamma^{-1}h$
- Equivalent to Gaussian BP

# Quadratic Minimization

$$f(x_1, ..., x_n) = \sum_i \frac{1}{2}\Gamma_{ii}x_i^2 - h_i x_i + \sum_{i>j} \Gamma_{ij} x_i x_j$$

- For min-sum algorithm, convergence to a unique estimate implies a solution to $\Gamma$x = h
- Corresponding graph has a node for each $x_i$ and an edge from i to j for each nonzero $\Gamma_{ij}$

# Covers of Quadratics

- Every k-cover of a quadratic on n variables is a quadratic on kn variables with

$$\widetilde{\Gamma} = \begin{pmatrix} \Gamma_{11}P_{11} & \cdots & \Gamma_{1n}P_{1n} \\ \vdots & \ddots & \vdots \\ \Gamma_{n1}P_{n1} & \cdots & \Gamma_{nn}P_{nn} \end{pmatrix}$$

$$\widetilde{h}_i = h_{\lceil i/k \rceil}$$

# Averaging

- We can scale solutions on a k-cover, $\widetilde{\Gamma} x' = \widetilde{h}$, down to solutions on the original problem by averaging copies:

$$x_i = \sum_{j \text{ copy of } i} \frac{x'_j}{k}$$

- Easy to check that $\Gamma x = h$

# Covers of Quadratics

$$\Gamma = \begin{pmatrix} 1 & .6 & .6 \\ .6 & 1 & .6 \\ .6 & .6 & 1 \end{pmatrix}$$

$$\widetilde{\Gamma} = \begin{pmatrix} 1 & 0 & .6 & 0 & 0 & .6 \\ 0 & 1 & 0 & .6 & .6 & 0 \\ .6 & 0 & 1 & 0 & .6 & 0 \\ 0 & .6 & 0 & 1 & 0 & .6 \\ 0 & .6 & .6 & 0 & 1 & 0 \\ .6 & 0 & 0 & .6 & 0 & 1 \end{pmatrix}$$

Positive definite

2-cover is not positive definite

# Sufficient Conditions

- Scaled diagonal dominance is sufficient for convergence [Moallemi et al. 06]

- Walk-summability is sufficient for convergence [Malioutov et al. 06]

- Min-sum converges for some matrices satisfying neither of these conditions

# Sufficient Conditions

- A symmetric matrix is scaled diagonally dominant if there exists w > 0 such that for each row i

$$\sum_j |\Gamma_{ij}| w_j < |\Gamma_{ii}| w_i$$

- A matrix is walk-summable if $\rho(|I-\Gamma|) < 1$

# Quadratic Minimization

- Want all covers of our matrix to be positive definite
  - Scaled diagonal dominance implies all covers scaled diagonally dominant
  - Scaled diagonal dominance implies positive definite

# Quadratic Minimization

- **Theorem** The following are equivalent
  - $\Gamma$ is walk-summable
  - $\Gamma$ is scaled diagonally dominant
  - Every cover of $\Gamma$ is positive definite
  - Every 2-cover of $\Gamma$ is positive definite
- Graph covers provide a "natural" sufficient condition for convergence

# Periodic Behavior

$$p(x_1, x_2, x_3) = \frac{1}{2}x^T \begin{pmatrix} 1 & .5 & .5 \\ .5 & 1 & .5 \\ .5 & .5 & 1 \end{pmatrix} x - \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}^T x$$

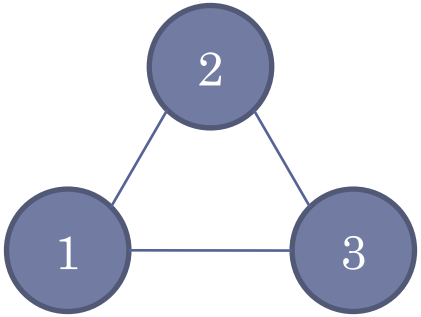
- Estimates alternate between 0 and 1 at each time step
- Beliefs are converging to constant functions
- Can we use covers to understand this?

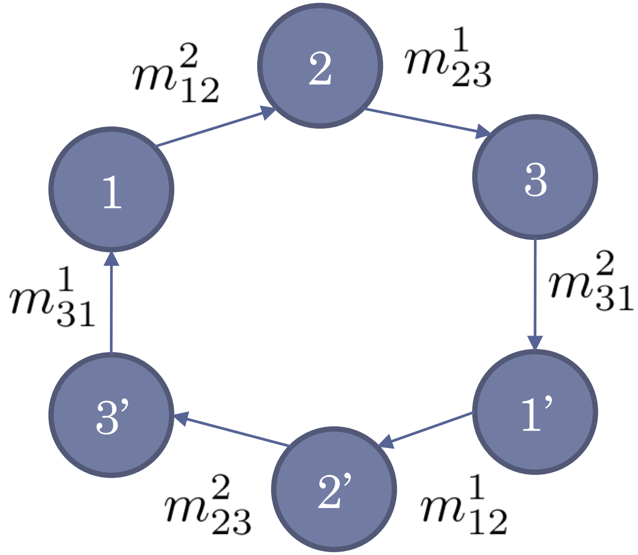
# Periodicity in Pairwise MRFs

- What other properties of min-sum do covers capture?
- Messages converge k-periodically if for each s the sequence of messages $m^{kt+s}$ converges as t tends to infinity
- Two easy cases to analyze periodicity:
  - Single cycles
  - 2-periodic messages

# Single Cycles

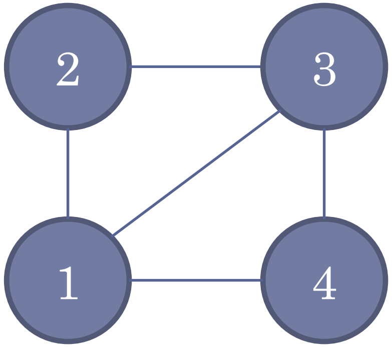
- We can fix k-periodic behavior by imagining a new message passing scheme on a k-cover

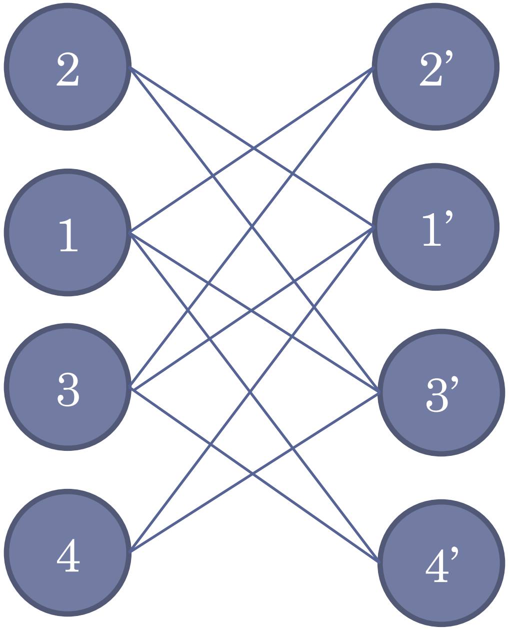


# 2-Periodicity

- For an arbitrary pairwise MRF if the messages converge 2-periodically we can fix the periodicity by looking at a bipartite 2-cover



# Periodicity

$$p(x_1, x_2, x_3) = \frac{1}{2}x^T \begin{pmatrix} 1 & .5 & .5 \\ .5 & 1 & .5 \\ .5 & .5 & 1 \end{pmatrix} x - \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}^T x$$

- Estimates alternate between 0 and 1 at each time step
- Beliefs are converging to constant functions so we cannot extract a unique assignment

# Periodicity and Global Optimality

- For the quadratic minimization problem, we can improve Wainwright's local optimality result to handle periodicity

  - We don't require that the fixed point beliefs converge to a unique estimate; only that the messages converge periodically

  - Use averaging procedure to obtain fixed point messages

# Conclusions

- Graph covers explain what makes certain problem instances harder than others
- Graph covers may be useful for deriving sufficient conditions for the convergence of BP in many application areas
- In many cases, periodic behavior can be understood as fixed point behavior on covers

Questions?