Walk-Sums and Belief Propagation in Gaussian Graphical Models

by Malioutov, Johnson and Willsky

presented by Ali Osman Ulusoy

Multivariate Gaussians

- Standard form: $p(x) \propto \exp\{-\frac{1}{2}(x-\mu)^T P^{-1}(x-\mu)\}$
  where P is symmetric and $P \succ 0$.

- Information form: $p(x) \propto \exp\{-\frac{1}{2}x^T Jx + h^T x\}$
  where J is symmetric and $J \succ 0$.

$$\mu = J^{-1}h \quad \text{and} \quad P = J^{-1}$$

J : Information matrix

h : Potential vector

Multivariate Gaussians

- Rescale the variables so that $J_{ii} = 1$
- Partial correlation coefficients

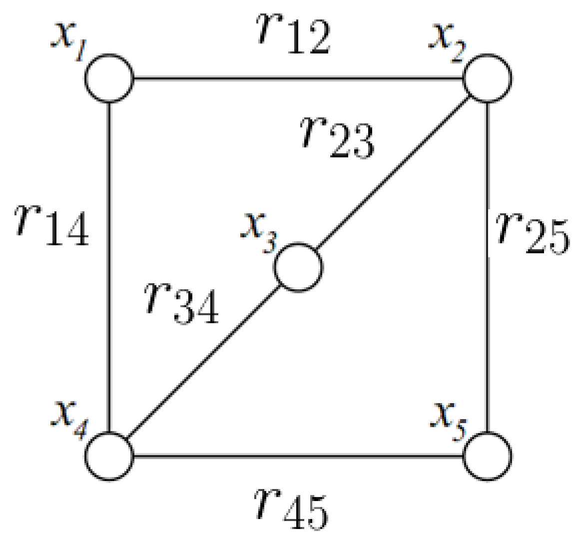
$$r_{ij} \triangleq \frac{\mathrm{cov}(x_i; x_j | x_{V \setminus ij})}{\sqrt{\mathrm{var}(x_i | x_{V \setminus ij}) \mathrm{var}(x_j | x_{V \setminus ij})}} = -\frac{J_{ij}}{\sqrt{J_{ii} J_{jj}}} = -J_{ij}$$

- Define R such that $R_{ii} = 0 \text{ and } R_{ij} = r_{ij}$

$$J = I - R$$

Gaussian MRFs

- $p(x_1, x_2, x_3, x_4, x_5)$ can be written as G = (V,E) where each component x_i is a node and r_{ij} are edge weights.

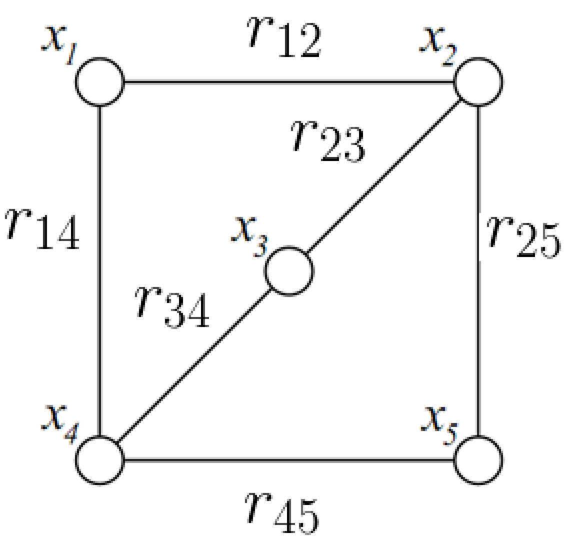


Sudderth 2002

Gaussian MRFs

- J encodes pairwise Markov independencies:

$$J_{ij} = 0 \quad \text{iff} \quad \{i, j\} \notin E$$



$$J = \begin{bmatrix} 1 & -r_{12} & 0 & -r_{14} & 0 \\ & 1 & -r_{23} & 0 & -r_{25} \\ & & 1 & -r_{34} & 0 \\ & & & 1 & -r_{45} \\ & & & & 1 \end{bmatrix}$$

Sudderth 2002

Gaussian MRFs

- Hammersley-Clifford theorem:

$p(x) \propto \exp\{-\frac{1}{2}x^T J x + h^T x\}$ can be written as

$$p(x) \propto \prod_{i \in V} \psi_i(x_i) \prod_{\{i,j\} \in E} \psi_{ij}(x_i, x_j)$$

$$\psi_i(x_i) = \exp\{-\frac{1}{2}J_{ii}x_i^2 + h_i x_i\}$$
$$\psi_{ij}(x_i, x_j) = \exp\{-x_i J_{ij} x_j\}$$

Problem

- Given the model (J_V, h_V), we want to perform variable elimination/marginalization

$$U = V \backslash i \qquad p_U(x_U) = \int_{x_i} p(\mathbf{x}) dx_i = N(\mu_U, P_U)$$

$$p(x_i) = \int_{x_{V \backslash i}} p(\mathbf{x}) dx_{V \backslash i} = N(\mu_i, \sigma_i)$$

Variable elimination in trees

For the acyclic Gaussian case, $p_U(x_U)$ is obtained by

$$\hat{J}_U = J_{U,U} - J_{U,i}J_{ii}^{-1}J_{i,U} \quad \text{and} \quad \hat{h}_U = h_U - J_{U,i}J_{ii}^{-1}h_i$$

$$p_U(x_U) = N(\hat{J}_U, \hat{h}_U)$$

How to compute all marginals efficiently?

Belief Propagation

Perform Sum-Product/BP to obtain marginals at each node.

$$m_{i \to j}(x_j) = \int \psi_{ij}(x_i, x_j) \psi_i(x_i) \prod_{k \in \mathcal{N}(i) \setminus j} m_{k \to i}(x_i) \; dx_i$$

$$p_i(x_i) \propto \psi_i(x_i) \prod_{k \in \mathcal{N}(i)} m_{k \to i}(x_i)$$

Cyclic case

- Try Loopy BP

$$m_{i\to j}^{(n)}(x_j) = \int \psi_{ij}(x_i, x_j)\psi_i(x_i) \prod_{k\in\mathcal{N}(i)\setminus j} m_{k\to i}^{(n-1)}(x_i)\ dx_i$$

- Not guaranteed to converge, energy function has multiple fixed points.
- Can we exploit Gaussian structure to guarantee convergence and correctness?

Walk-Sums 101

- Walk $w = (w_0, w_1, \ldots, w_l)$ $w_k \in V$

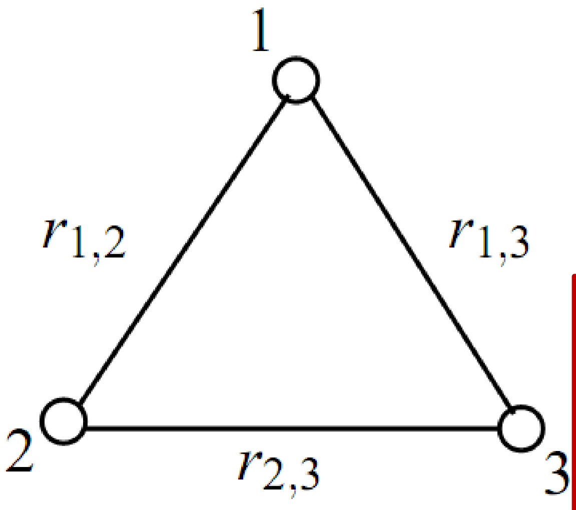
  $\{w_k, w_{k+1}\} \in E$

- Weight of a walk w $\phi(w) = \prod_{k=1}^{l(w)} r_{w_{k-1}, w_k}$

- A set of walks W

$$\phi(\mathcal{W}) = \sum_{w \in \mathcal{W}} \phi(w) \qquad \phi_h(\mathcal{W}) = \sum_{w \in \mathcal{W}} h_{w_0} \phi(w)$$

Walk-Sum Example



$$w = (1,2,3) \quad \phi(w) = r_{1,2} r_{2,3}$$

$$\mathcal{W}(1 \to 1): \quad \{(1),(1,2,1),(1,3,1),(1,3,2,1),(1,2,1,2,1),\ldots\}$$

$$\phi(1 \to 1) = 1 + r_{1,2} r_{2,1} + r_{1,3} r_{3,1} \ldots.$$

$$\mathcal{W}(* \to 1): \quad \{(1),(2,1),(3,1),(2,3,1),(1,3,1),\ldots\}$$

$$\phi_h(* \to 1) = h_1 + h_2 r_{2,1} + h_3 r_{3,1} + h_2 r_{2,3} r_{3,1} + \ldots.$$

Neumann Series

$$J = I - R$$

Maximum absolute value of the eigenvalues.

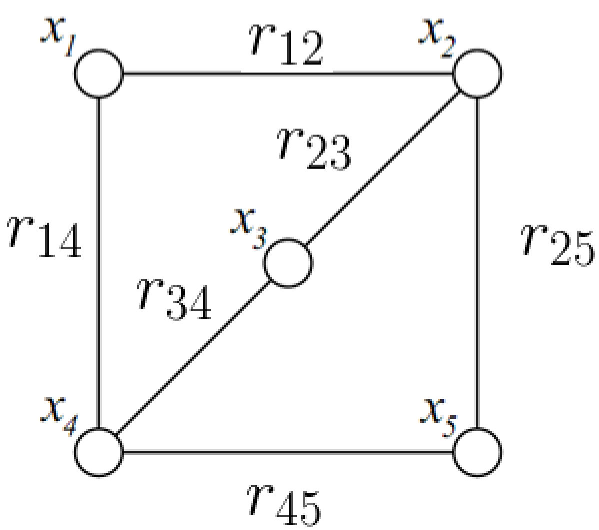
$$P = J^{-1} = (I - R)^{-1} = \sum_{k=0}^{\infty} R^k, \quad \text{for} \quad \rho(R) < 1$$

- $(R^l)_{ij}$ can be interpreted as sum of walks from i to j of length l

$$(R^l)_{ij} = \sum_{w_1,\dots,w_{l-1}} r_{i,w_1} r_{w_1,w_2} \dots r_{w_{l-1},j} = \sum_{w:i \xrightarrow{l} j} \phi(w)$$

Neumann Series

Compute $(R^2)_{13}$



$$R = \begin{bmatrix} 0 & r_{12} & 0 & r_{14} & 0 \\ r_{21} & 0 & r_{23} & 0 & r_{25} \\ 0 & r_{32} & 0 & r_{34} & 0 \\ r_{41} & 0 & r_{43} & 0 & r_{45} \\ 0 & r_{52} & 0 & r_{54} & 0 \end{bmatrix}$$

$$(R^2)_{13} = [0 \;\; r_{12} \;\; 0 \;\; r_{14} \;\; 0] \times \begin{bmatrix} 0 \\ r_{23} \\ 0 \\ r_{43} \\ 0 \end{bmatrix} = r_{12}r_{23} + r_{14}r_{43}$$

Walk-Sums for Inference

$$P = J^{-1} = (I - R)^{-1} = \sum_{k=0}^{\infty} R^k, \quad \text{for} \quad \rho(R) < 1$$

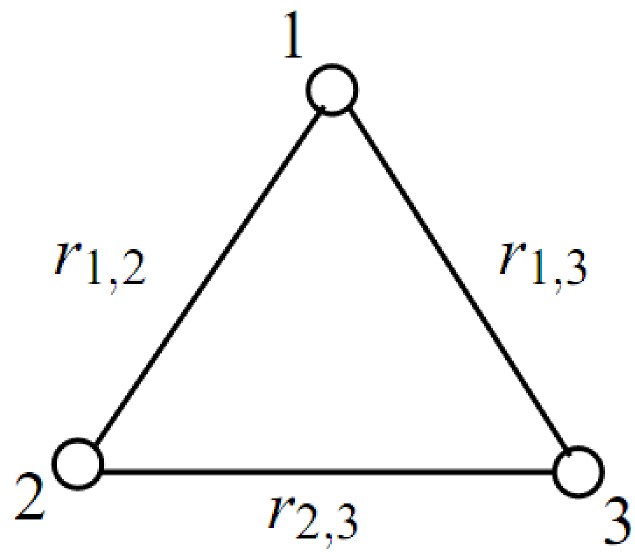
- Computing the marginals

$$P_{ii} = \phi(i \to i) = \frac{1}{1 - \alpha_i} \qquad \alpha_i = \phi(i \xrightarrow{\backslash i} i)$$

$$\mu_i = \phi_h(* \to i) = \frac{h_i + \beta_i}{1 - \alpha_i} \qquad \beta_i = \phi_h(* \xrightarrow{\backslash i} i)$$

Walk-Sums for Inference

- Example



$$P_{1,1} = \phi(1 \rightarrow 1) = 1 + r_{1,2}r_{2,1} + r_{1,3}r_{3,1} + r_{1,2}r_{2,3}r_{3,1} + \ldots.$$

$$\mu_1 = \phi_h(* \rightarrow 1) = h_1 + h_2 r_{2,1} + h_3 r_{3,1} + h_2 r_{2,3} r_{3,1} + \ldots.$$

Sums are guaranteed to converge!

Walk-Summability

- If the sum is well-defined (converges absolutely) then the model is walk summable.
- Equivalent conditions for walk-summability

(i) $\sum_{w:i \to j} |\phi(w)|$ *converges for all* $i, j \in V$.

(ii) $\sum_l \bar{R}^l$ *converges*.

(iii) $\rho(\bar{R}) < 1$.

(iv) $I - \bar{R} \succ 0$.

Define $\bar{R}$ where $\bar{R}_{ij} = |R_{ij}|$

$$P = J^{-1} = (I - R)^{-1}$$

$$= \sum_{k=0}^{\infty} R^k \quad \text{for} \quad \rho(R) < 1$$

Walk-Summable models

- Attractive models: If for all i,j $R \geq 0$
  - Attractive models are walk-summable. Proof follows directly from condition (iv).

- Non-frustrated models: If the graph doesn't contain cycles with an odd number of negative edge weights.
  - Non-frustrated models are walk-summable.
  - Trees don't contain any cycles so they are also walk-summable.

Walk-Summable models

- Pairwise-Normalizable: If for every edge $e \in E$,

$$J_e \succ 0$$

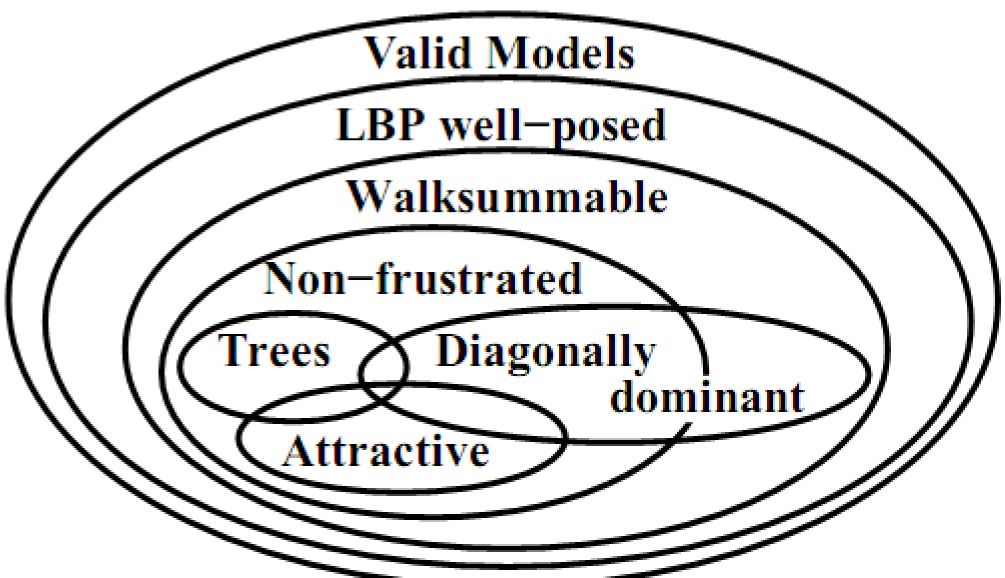
- Walk-Summability is **equivalent** to pairwise normalizability.

- Diagonally dominant: $\sum_{j \neq i} |J_{ij}| < J_{ii}$

These models are pairwise-normalizable and hence walk-summable.

Walk-Summable models

- Walk-summable/pairwise normalizable models include trees, attractive, non-frustrated and diagonally dominant models.



Walk-Sum interpretation of BP

- Walk-Sum computation on trees is equivalent to running BP.
- This framework can be used to analyze loopy BP behavior in cyclic graphs:
  - Loopy BP is equivalent to exact inference on the computation tree.
  - Then, Loopy BP is equivalent to walk sums in the computation tree.
  - Analyze the difference between walk sums in the computation tree and walk sums in the original graph.

Walk-Sums and BP on trees

- Calculating variances with Walk-Sums:

$$P_{jj} = \frac{1}{1 - \alpha_j}$$

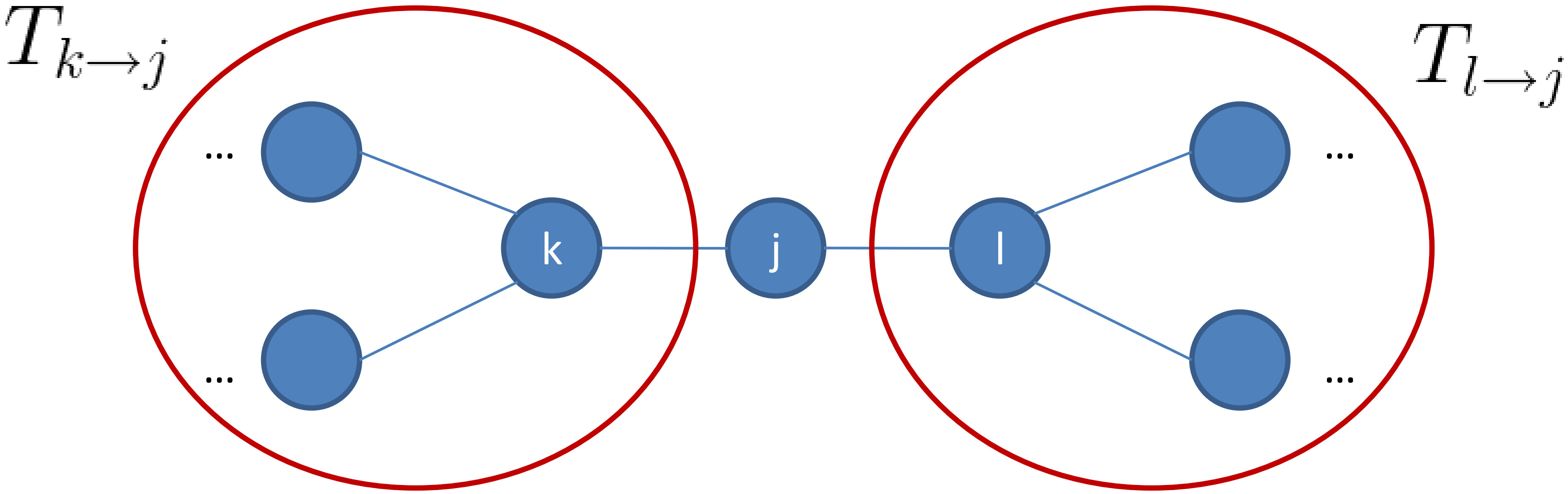
$$\alpha_j = \phi(j \xrightarrow{\setminus j} j) = \sum_{i \in \mathcal{N}(j)} \phi(j \xrightarrow{\setminus j} j \mid T_{i \to j}) \triangleq \sum_{i \in \mathcal{N}(j)} \alpha_{i \to j}$$

- Belief equation at node j

$$b(x_j) \propto \psi_j(x_j) \prod_{k \in N(j)} m_{k \to j}(x_j)$$

Walk-Sums and BP on trees

$$\alpha_j = \phi(j \xrightarrow{\backslash j} j) = \sum_{i \in \mathcal{N}(j)} \phi(j \xrightarrow{\backslash j} j \mid T_{i \to j}) \triangleq \sum_{i \in \mathcal{N}(j)} \alpha_{i \to j}$$

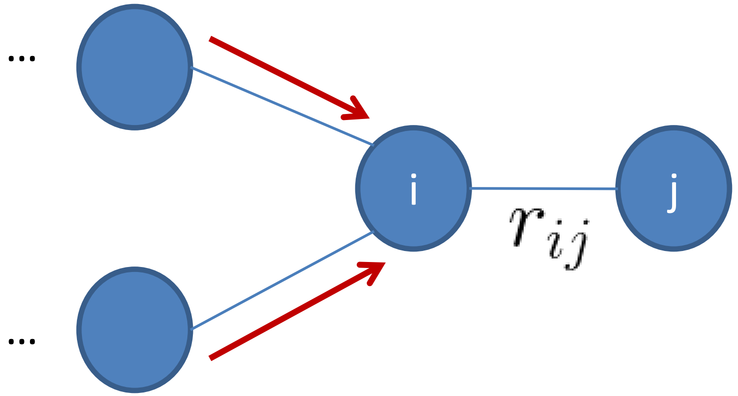


$$\alpha_j = \alpha_{l \to j} + \alpha_{k \to j}$$

Walk-Sums and BP on trees

- How about $\alpha_{i \to j}$? How do they correspond to messages in BP?

$$\alpha_{i \to j} = r_{ij}^2 \frac{1}{1 - \sum_{k \in \mathcal{N}(i) \setminus j} \alpha_{k \to i}}$$

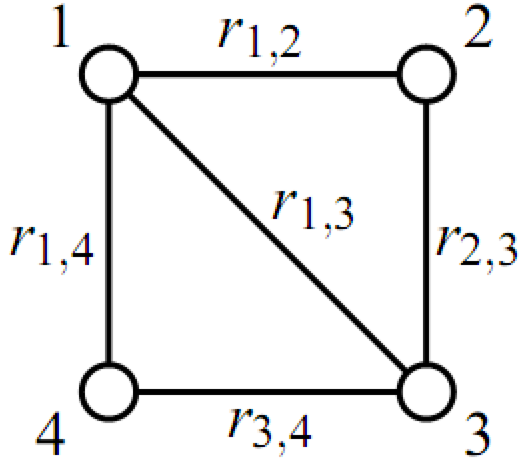


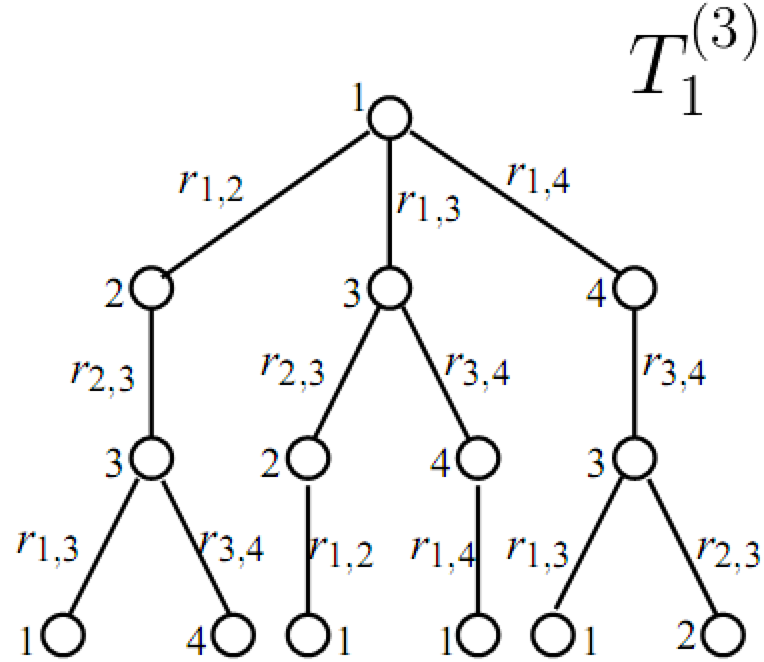
How about loopy BP?

- Can we use the walk-sum framework to understand the convergence and correctness of loopy BP in Gaussian models?
  - We will do this by comparing the walk-sums in the original graph to the loopy BP, which is equivalent to walk-sums in the computation tree.
  - Are they the same?

Computation tree

- Running Loopy BP on the original graph is equivalent to running exact inference on the computation tree $T_i^{(n)}$, hence doing walk-sums.





Loopy BP in Walk-Summable Models

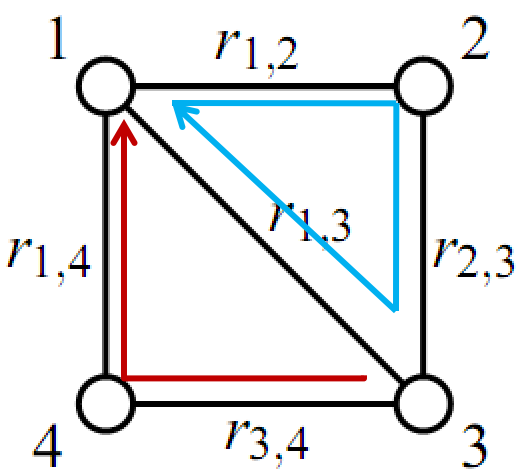
- After n iterations, the estimates for node 0 (root of the tree) are

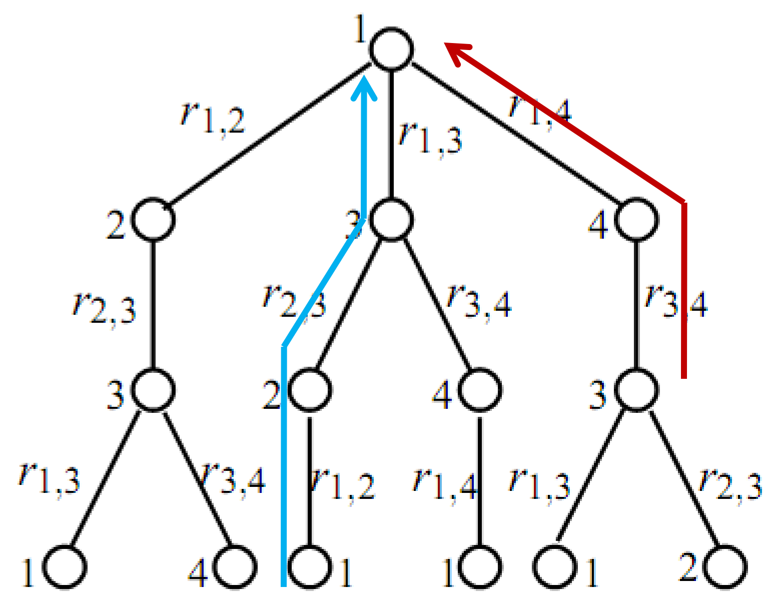
$$P_0(T_i^{(n)}) = \phi(0 \rightarrow 0 \mid T_i^{(n)})$$
$$\mu_0(T_i^{(n)}) = \phi_h(* \rightarrow 0 \mid T_i^{(n)})$$

- Assuming loopy BP has converged, are the variance and mean estimates correct?

Estimated means

- Lemma: There is a 1-1 relationship between walks in the original graph and walks in the computation tree.

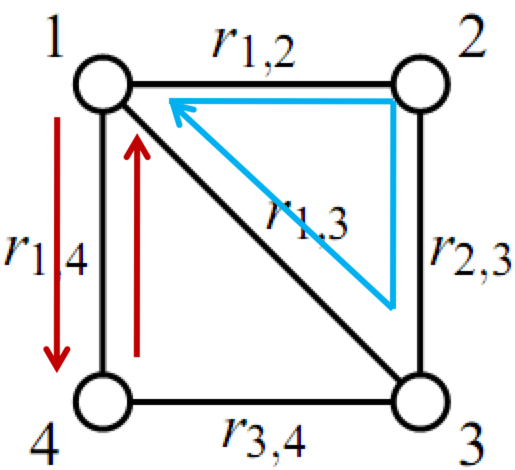


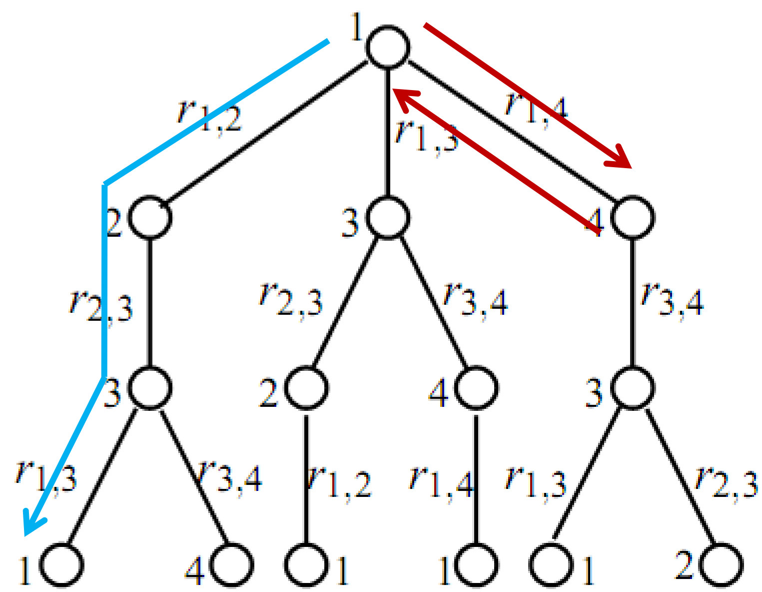


- Upon convergence, the estimated means are **correct**!

Estimated variances

- Lemma: Loopy BP variance estimate is a sum of *backtracking* self-return walks, a subset of all self-return walks.





- Upon convergence, the estimated variances are **incorrect**!

Convergence

- All walks in $T_i^{(n)}$ are subsets of walks in the original graph.
- We have already shown that latter converges so the former must also converge!

- Loopy BP in walk-summable models will **always converge**!

Summary

- Introduced Walk-Sum framework.
- Shown that many non-trivial classes of models are walk-summable.
- Presented a Walk-Sum interpretation of BP.
- Shown that Loopy BP will converge in walk-summable models and upon convergence, the means will be correct but variances in general will not.