Inference over discrete FGs with cycles

Introduction to Graphical Models and Inference for Communications

UC3M

March 3, 2018

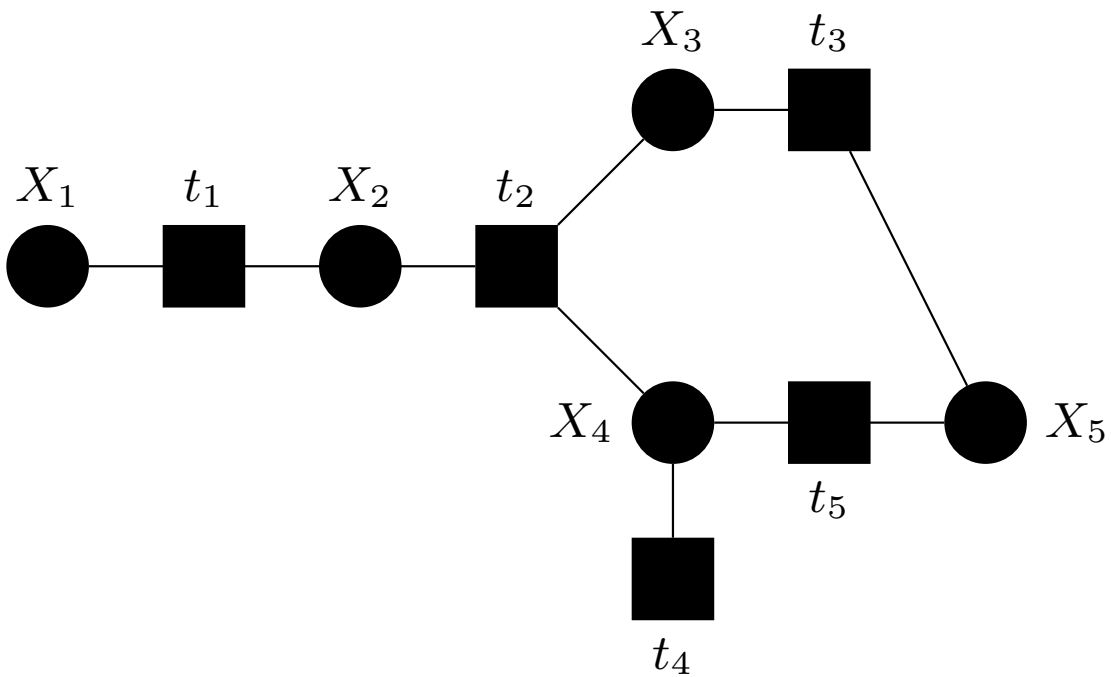
uc3m

Today

- We want to perform inference over discrete probability distributions that are represented by Factor Graphs that are not trees, the graph has cycles!
- **Exact inference is in general $\mathcal{O}(|\mathcal{X}|^n)$.**
- Approximate Inference using BP!
  - **Very fast :) ... poor accuracy in general :(**
  - **In some scenarios the BP accuracy is remarkable!**

Index

$$p_{\boldsymbol{X}}(\boldsymbol{x}) = \frac{1}{Z} t_1(x_1, x_2) t_2(x_2, x_3, x_4) t_3(x_3, x_5) t_4(x_4) t_5(x_4, x_5), \qquad \boldsymbol{x} \in \mathcal{X}^5$$

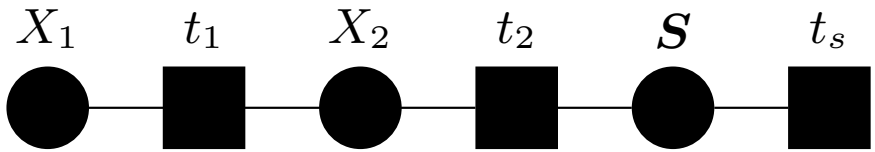


By defining $\boldsymbol{S} = \begin{bmatrix} X_3 & X_4 & X_5 \end{bmatrix}^T$, the same distribution factorizes as follows

$$p_{\boldsymbol{X}}(\boldsymbol{x}) = \frac{1}{Z} t_1(x_1, x_2) t_2(x_2, x_3, x_4) \underbrace{t_3(x_3, x_5) t_4(x_4) t_5(x_4, x_5)}_{t_s(\boldsymbol{s})}$$

$$p_{\boldsymbol{X}}(\boldsymbol{x}) = \frac{1}{Z} t_1(x_1, x_2) t_2(x_2, \boldsymbol{s}) t_s(\boldsymbol{s})$$

$$x_3 = \begin{bmatrix} 1 & 0 & 0 \end{bmatrix} \boldsymbol{s}, \qquad x_4 = \begin{bmatrix} 0 & 1 & 0 \end{bmatrix} \boldsymbol{s}, \qquad x_5 = \begin{bmatrix} 0 & 0 & 1 \end{bmatrix} \boldsymbol{s},$$



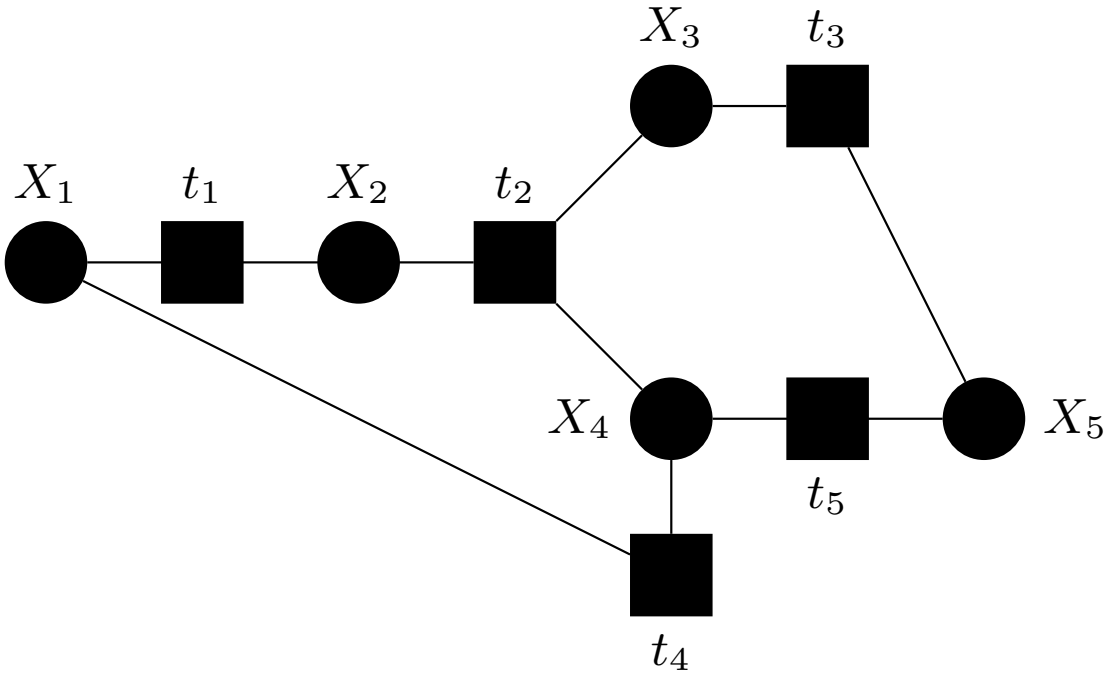
BP is exact!

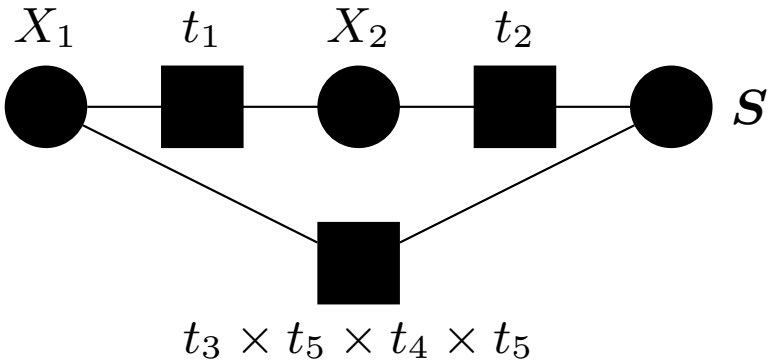
Increased complexity

Computing the message

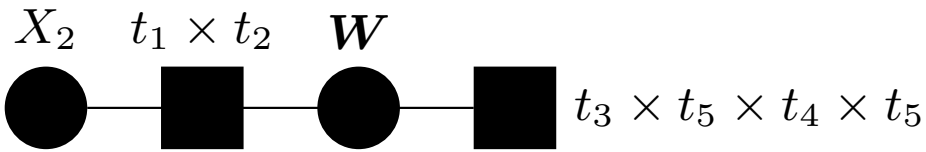
$$m_{t_2 \to x_2}(x_2) = \sum_{\boldsymbol{s}} t_j(x_2, \boldsymbol{s}) t_s(\boldsymbol{s})$$

requires $\mathcal{O}(|\mathcal{X}|^3)$ operations.

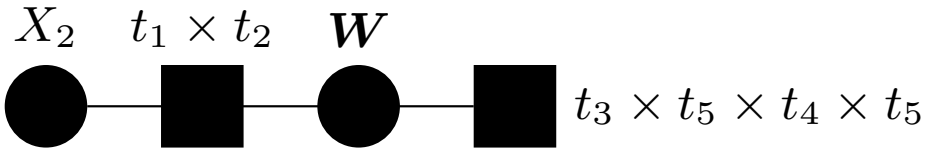




$$\boldsymbol{S} = [\ X_3 \quad X_4 \quad X_5\ ]^T$$



$$\boldsymbol{W} = [\ X_1 \quad X_3 \quad X_4 \quad X_5\ ]^T$$



$$\boldsymbol{W} = [\ X_1 \quad X_3 \quad X_4 \quad X_5\ ]^T$$

BP is exact!

Increased complexity

Computing the message

$$m_{t_1 \times t_2 \to x_2}(x_2) = \sum_{\boldsymbol{w}} t_1(\boldsymbol{w}, x_2) t_2(\boldsymbol{w}, x_2)$$

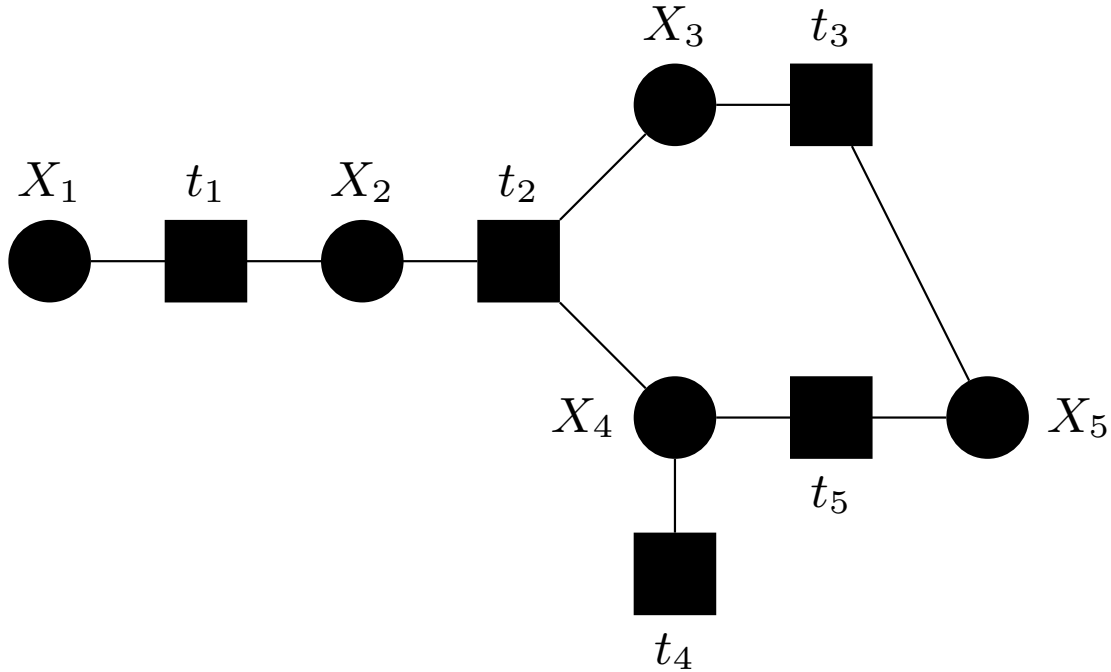
requires $\mathcal{O}(|\mathcal{X}|^4)$ operations.

For high-dimensional, high-density graphs $\to \mathcal{O}(|\mathcal{X}|^{\alpha n})$ complexity.

Cutset conditioning

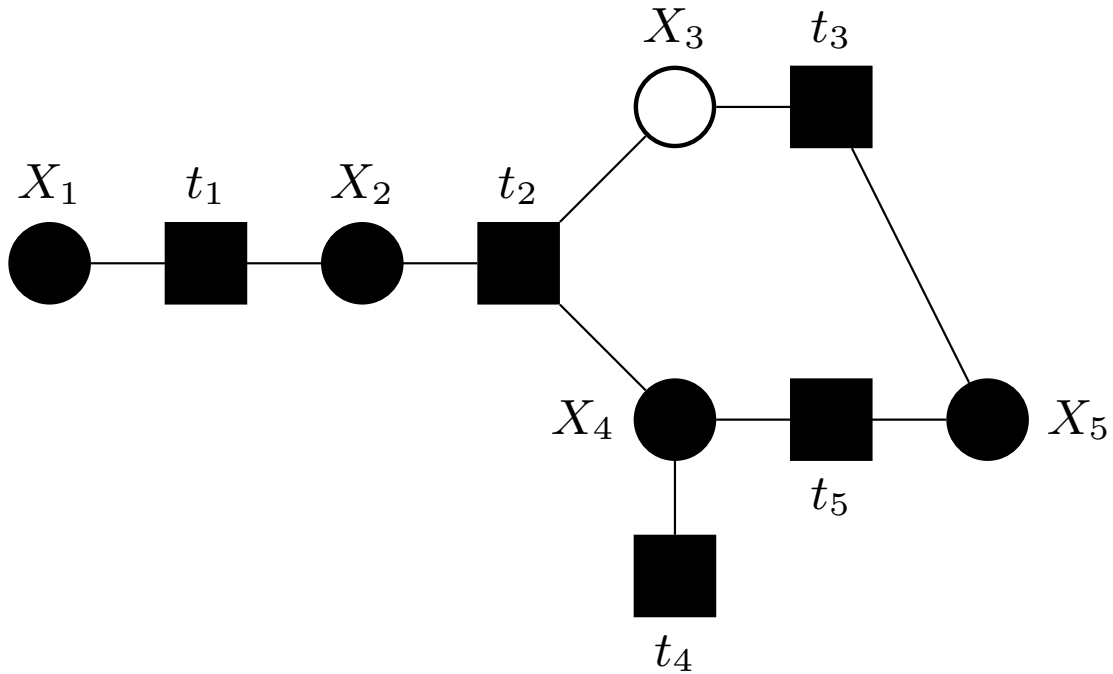
- An alternative to clustering that seeks to solve inference over a graph with cycles by working with cycle-free graphs.
- It can be used in conjuntion to clustering.
- For high-dimensional, high-density graphs $\rightarrow \mathcal{O}(|\mathcal{X}|^{\alpha n})$ complexity.

$$p_{\boldsymbol{X}}(\boldsymbol{x}) = \frac{1}{Z} t_1(x_1, x_2) t_2(x_2, x_3, x_4) t_3(x_3, x_5) t_4(x_4) t_5(x_4, x_5), \qquad \boldsymbol{x} \in \mathcal{X}^5$$

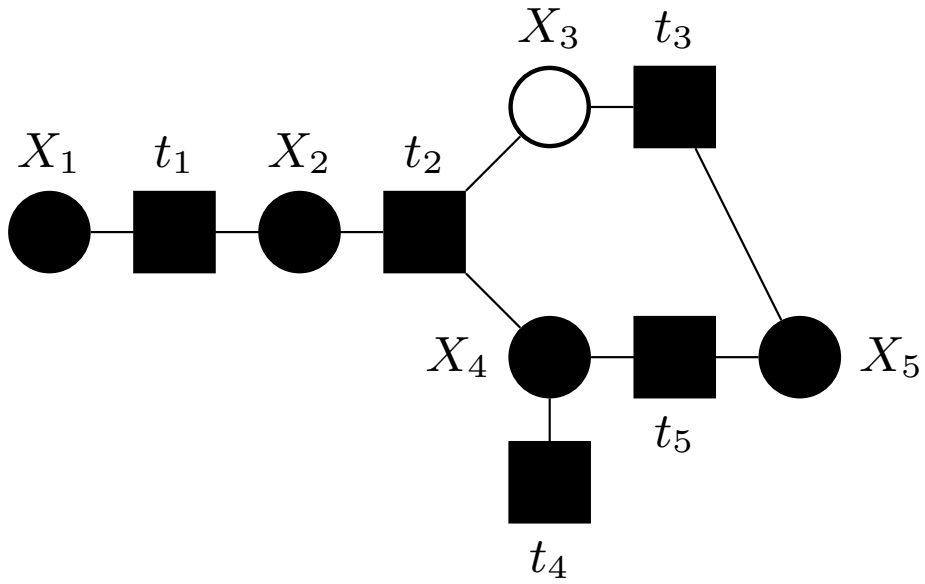


$$\mathcal{X} = \{\mathsf{a}_1, \mathsf{a}_2, \ldots, \mathsf{a}_{|\mathcal{X}|}\}$$

$$p_{\boldsymbol{X}}(x_1, x_2, x_3 = \mathtt{a}_1, x_4, x_5)$$
$$= \frac{1}{Z} t_1(x_1, x_2) t_2(x_2, x_3 = \mathtt{a}_1, x_4) t_3(x_3 = \mathtt{a}_1, x_5) t_4(x_4) t_5(x_4, x_5), \qquad \boldsymbol{x} \in \mathcal{X}^5$$



$p_{\boldsymbol{X}}(x_1, x_2, x_3 = \mathtt{a}_1, x_4, x_5)$ maps over a cycle-free FG!



Using BP, we can compute

$$p_{X_3}(x_3 = \mathtt{a}_1)$$
$$p_{X_1,X_3}(x_1, x_3 = \mathtt{a}_1)$$
$$p_{X_2,X_3}(x_2, x_3 = \mathtt{a}_1)$$
$$p_{X_4,X_3}(x_4, x_3 = \mathtt{a}_1)$$
$$p_{X_5,X_3}(x_5, x_3 = \mathtt{a}_1)$$

at cost $\mathcal{O}(n|\mathcal{X}|^3)$.

We can proceed similarly to compute

$$p_{X_3}(x_3 = \mathtt{a}_j)$$
$$p_{X_i,X_3}(x_i, x_3 = \mathtt{a}_j)$$

for all $a_j \in \mathcal{X}$ and $i = 1, 2, 4, 5$. The marginal for X_3 is already computed, and the rest of marginals...

$$p_{X_i}(x_i) = \sum_{x_3 \in \mathcal{X}} p_{X_i,X_3}(x_i, x_3)$$

For a single cycle, the overall complexity is ...

$$\mathcal{O}(n|\mathcal{X}|^3 \times |\mathcal{X}|)$$

For a high-dimensional, high-density graph the # of cycles is proportional to n

$$\mathcal{O}(n|\mathcal{X}|^d \times |\mathcal{X}|^{\alpha n})$$

Index

Just some words about the Junction Tree Algorithm (JTA)

- Systematic procedure to perform **exact inference** over arbitrary graphs.
- Construct tree-graph representations of our graph via clustering.
- Given the FG of our problem, we construct a **cluster graph** with the properties of a **junction tree.**
- Run a generalization of BP over cluster graphs. Exact for a junction tree!

In a great deal of interesting applications the use of the JTA algorithm would result in clusters that are prohibitively large. $\mathcal{O}(|\mathcal{X}|^{\alpha n})$ complexity.

JTA complexity is typically affordable for highly regular graphs!

- Hidden Markov model, state-space models.
- Lattice-models (widely used in image statistical modelling ...)

Take a look of Chapter 6, David Barber's book!.

Index

The need for approximations

- For many problems of practical interest, it will not be feasible to use exact inference (JTA).
- We need to exploit effective approximation methods.
  - **Deterministic approaches.** Variational methods.
  - **Monte Carlo methods**. Based on stochastic numerical sampling.
- Simplest approximate inference method: **loopy-BP**.

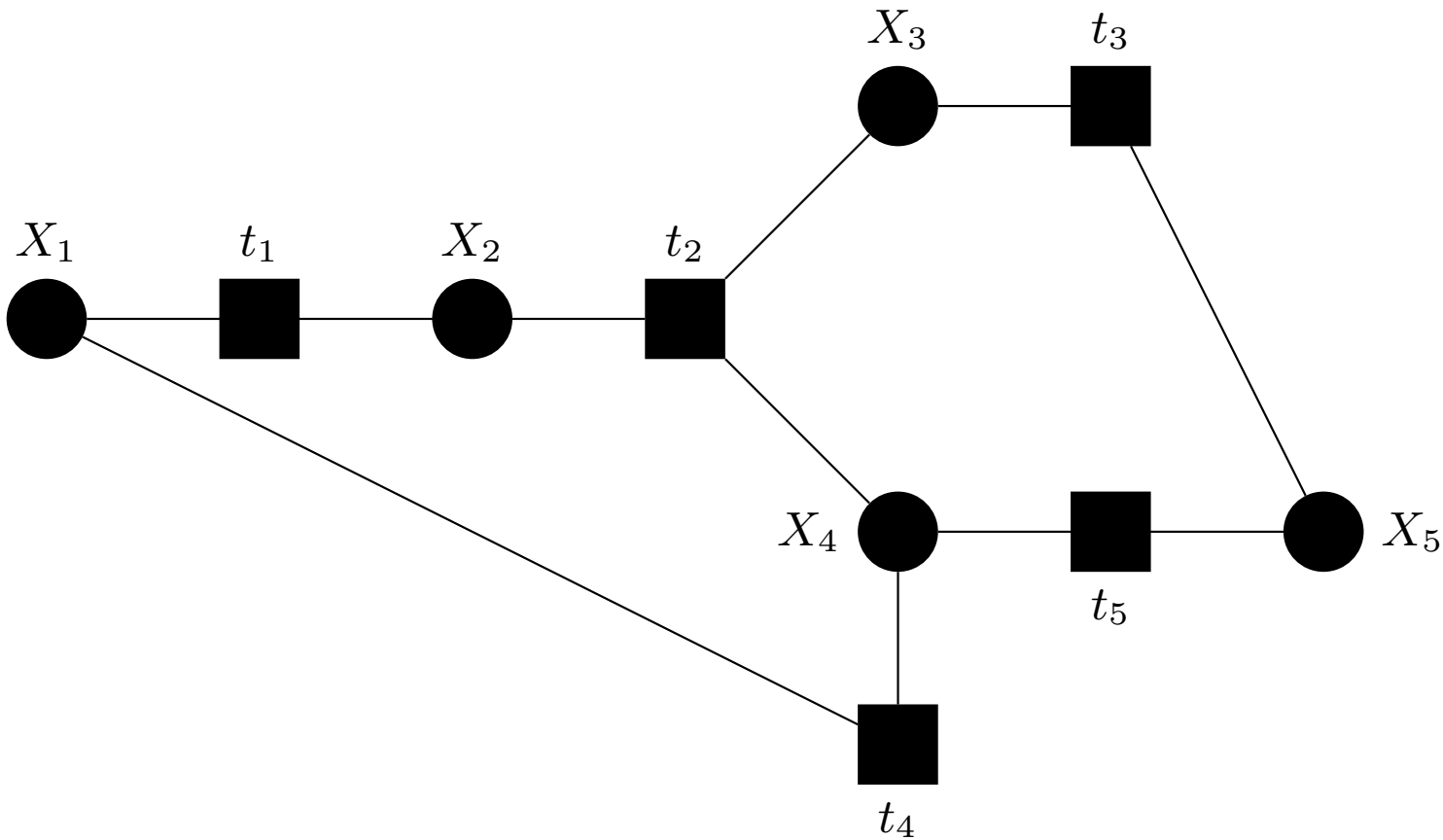
Loopy Belief Propagation

- Apply the exact same BP rules, ignoring the fact that the FG may contain cycles.
- Not a problem because the BP message passing rules are purely local.
- An "iterative" algorithm with no natural termination will result. Messages passed multiple times on a given edge.
- **Convergence is not guaranteed**. For some models, BP converges for some others not.

Global initialization

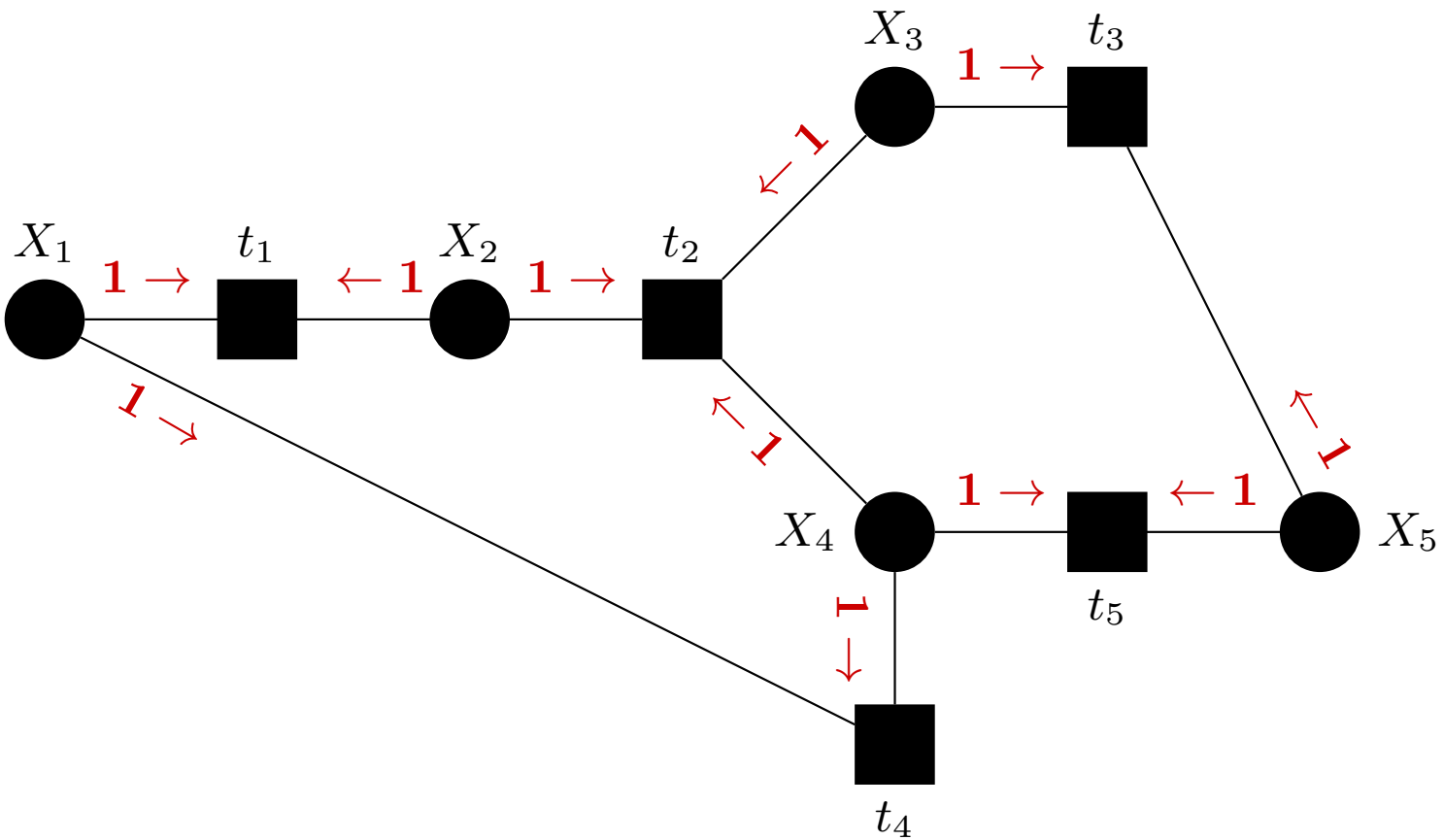
We set all the initial messages from variables to 1.

$$\text{For all variable nodes:} \quad m_{x_i \to t_j}(x_i) = 1 \quad x_i \in \mathcal{X}$$


X_3
t_3
X_1
t_1
X_2
t_2
X_4
X_5
t_5
t_4

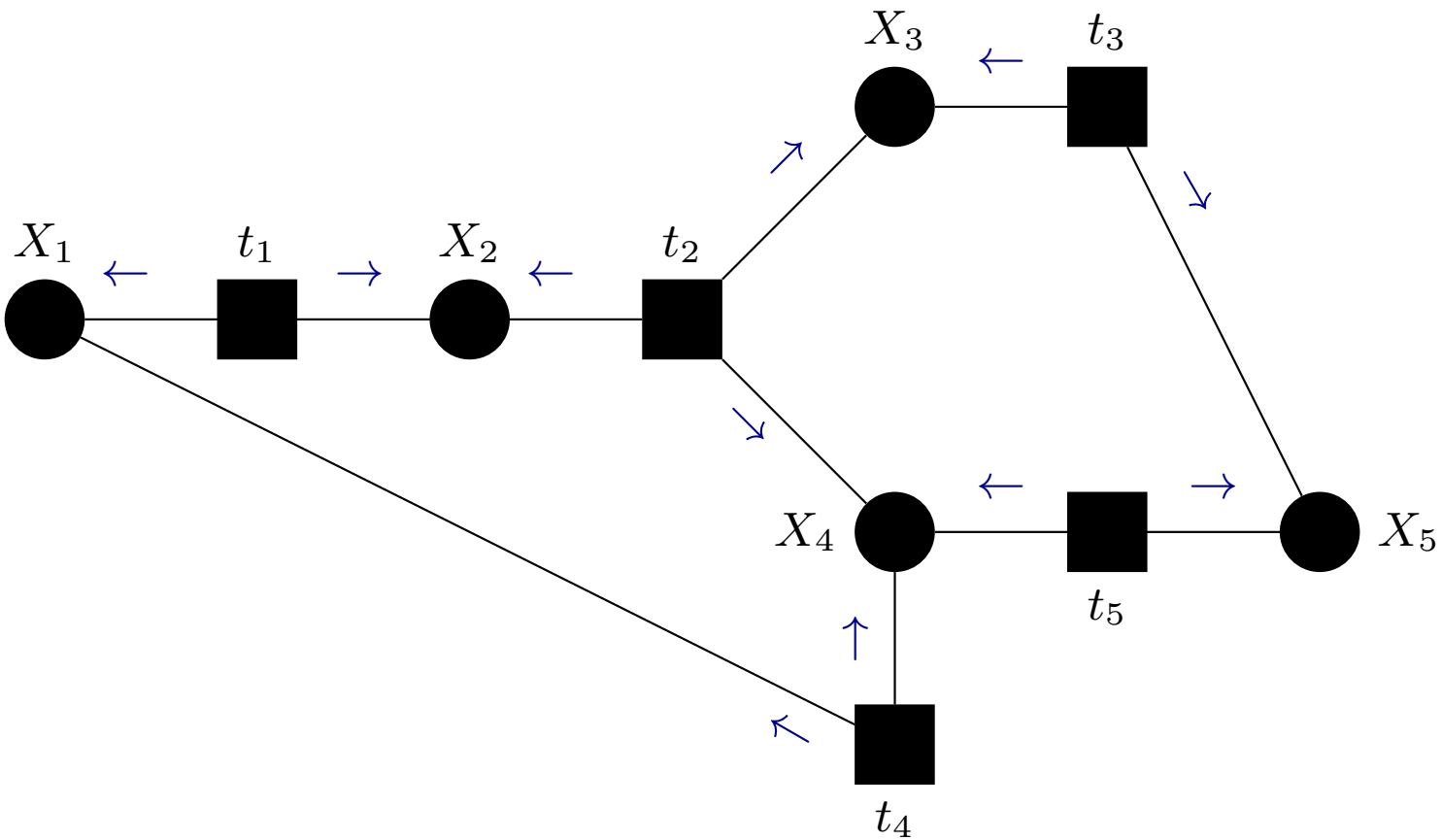
Global Initialization

- **1 →**: all-1s message.



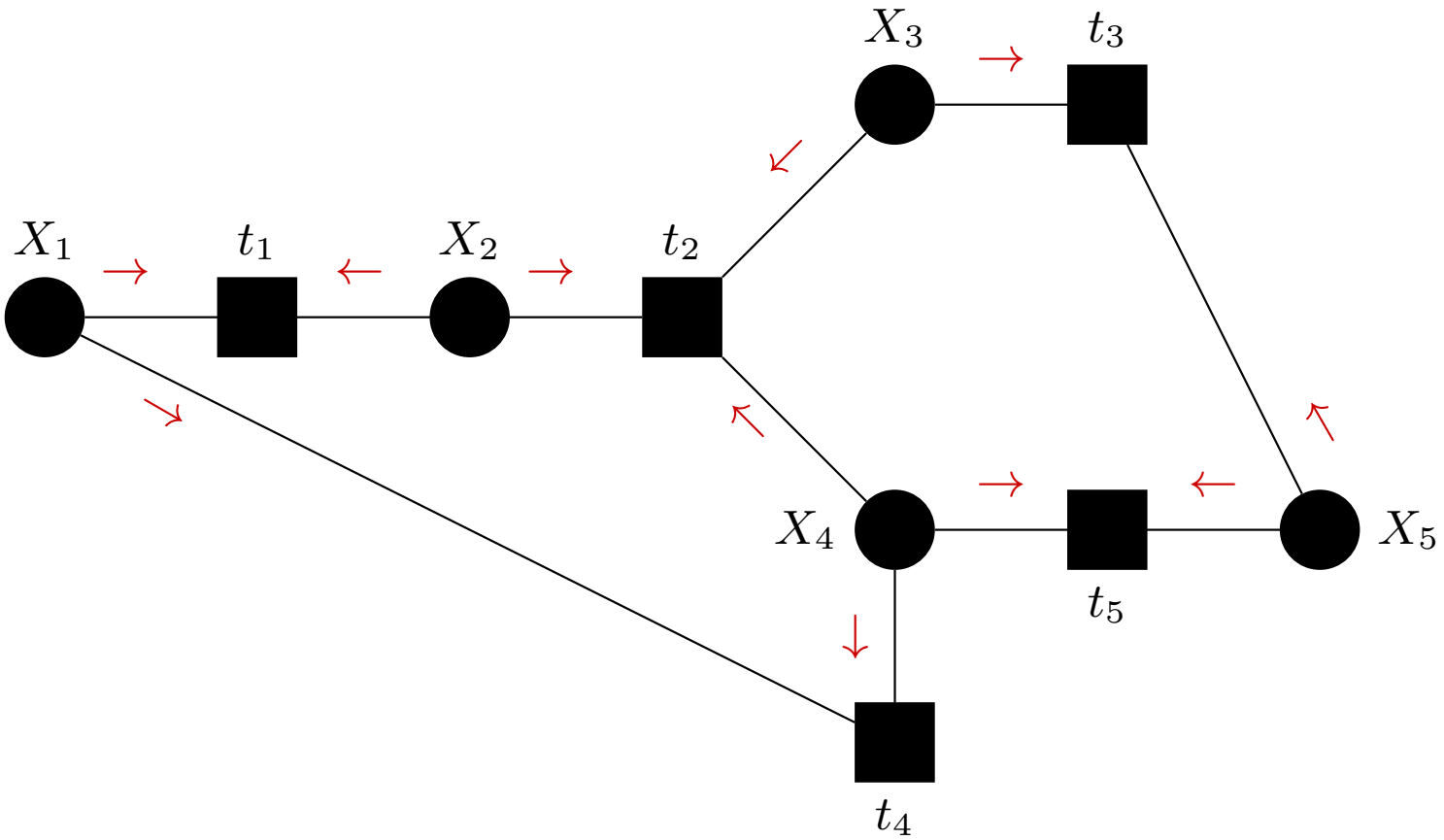
Flooding scheme (I)

- All factor nodes recompute a message for each neighbor.



Flooding scheme (II)

- All variable nodes recompute a message for each neighbor.



There is no guarantee that this update rules converge to a fixed point. In practice, all schedules are finite, either by running the algorithm for a finite-number of iterations or by another termination condition.

Approximated marginals

- Upon finalization, marginals are estimated using the product of the most recently received incoming messages to each variable node.

$$p_{X_i}(x_i) \approx \frac{1}{Z_i} \prod_{k \in \mathtt{Ne}(X_i)} m_{t_k \to x_i}(x_i),$$

- In some applications, loopy BP can give poor results, where in other applications it has proven to be very effective.

- **LDPC decoding**.
- **Signal reconstruction in compressed sensing applications**.
- **Symbol detection in massive MIMO scenarios**.
- ...

Analysis of Loopy-BP

Several works have yielded insight into the dynamics and convergence properties of loopy BP:

Weiss, Y. (1997). Belief propagation and revision in networks with loops. Technical report, Cambridge, MA, USA.

- For graphs with a **single-cycle**, it has been proven that the loopy BP always converges to a fixed point that can be analytically characterized.
- Further, if the hidden variables in the cycle are binary-valued, then a decisor based on the approximate marginals computed by loopy BP is equivalent to the MAP solution.

Analysis of Loopy-BP

Several works have yielded insight into the dynamics and convergence properties of loopy BP:

Weiss, Y. and Freeman, W. T. Correctness of belief propagation in gaussian graphical models of arbitrary topology. NIPS 1999.

- BP algorithm applied to a **Gaussian** pdf that maps over a cycle-free FG provides the exact marginals with a **complexity cubic in the cluster size**.
- Sufficient conditions for convergence of BP when the algorithm converges it gives the **exact posterior means** for all graph topologies, not just cycle-free FGs.

Analysis of Loopy-BP

Several works have yielded insight into the dynamics and convergence properties of loopy BP:

Yedidia, J. S., Freeman, W. T., and Weiss, Y. Generalized belief propagation. NIPS 2001.

- Fixed points of the loopy belief propagation are actually **stationary points of the Bethe free energy** from statistical physics.
- The Bethe free energy provides a firm **theoretical basis of loopy belief propagation** and it served as a basis for more advanced methods, such as generalized belief propagation.