HAL open science

Spatial Triplet Markov Trees for auxiliary variational inference in Spatial Bayes Networks

Hugo Gangloff, Jean-Baptiste Courbot, Emmanuel Monfrini, Christophe Collet

To cite this version:

Hugo Gangloff, Jean-Baptiste Courbot, Emmanuel Monfrini, Christophe Collet. Spatial Triplet Markov Trees for auxiliary variational inference in Spatial Bayes Networks. SMTDA 2020: 6th international conference on Stochastic Modeling Techniques and Data Analysis, Jun 2020, Barcelone (online), Spain. pp.237-249. hal-03122810

HAL Id: hal-03122810

https://hal.science/hal-03122810v1

Submitted on 30 Jan 2023

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Spatial Triplet Markov Trees for Auxiliary Variational Inference in Spatial Bayes Networks

Hugo Gangloff[1,2]
Jean-Baptiste Courbot[3], Emmanuel Monfrini[4], Christophe Collet[1],

[1]Université de Strasbourg - CNRS ICube UMR 7357, Illkirch, France
[2]GEPROVAS, Strasbourg, France
[3]IRIMAS UR 7499 Université de Haute-Alsace, Mulhouse, France
[4]SAMOVAR, Télécom SudParis, IP Paris, Évry, France

Stochastic Modeling Techniques and Data Analysis, Virtual Conference, June, 2-5 2020

Outline

1. **Introduction**
2. **The probabilistic models**
   - Markov Tree network
   - Spatial Bayes Network
   - STMT network
3. **Variational Inference**
   - The method
   - Several VIs on a small example
   - Experimental results
4. **Conclusion & Perspectives**

Introduction

Bayesian Networks: main definitions

- Also known as acyclic Directed Graphical Models (DGM)
- Let $\mathcal{G} = (\mathcal{S}, \mathcal{E})$ be a directed graph:
  - Directed edges link *parents* vertices $\mathcal{P}(s)$ and son vertice $s \in \mathcal{S}$
  - Vertices without any parent are *root* vertices (denoted r)
- Random variables associated to each vertice
- Define *local conditional probabilities* $p(x_r)$ or $p(x_s|\boldsymbol{x}_{\mathcal{P}(s)})$
  $\rightarrow$ They define joint distribution on $\mathcal{G}$:

$$p(\boldsymbol{x}) = \prod_{r \in \text{roots}} p(x_r) \prod_{s \in \mathcal{S} \setminus \text{roots}} p(x_s|\boldsymbol{x}_{\mathcal{P}(s)}) \tag{1}$$

Inference basics

Inference tasks:

- compute marginals $p(x_s)$
- compute conditionals $p(x_s|\boldsymbol{x}_{p(s)})$
- compute a mode of the distribution
  (*e.g.* Maximum A Posteriori (MAP): $\hat{\boldsymbol{x}} = \mathrm{argmax}_{\boldsymbol{x}} p(\boldsymbol{x})$,
  Maximum Posterior Modes (MPM): $\hat{x_s} = \mathrm{argmax}_{x_s} p(x_s)$)

$$\downarrow$$

Direct computations in Bayes Networks:

- Forward Backward (Stratonovich 1965) / Belief Propagation (Pearl 1982)

Approximate computations in general DGM:

- Loopy Belief Propagation (Weiss 2000)
- MCMC: Gibbs sampling (Geman et al. 1984)
- **Variational Inference** (Jordan et al. 1999)

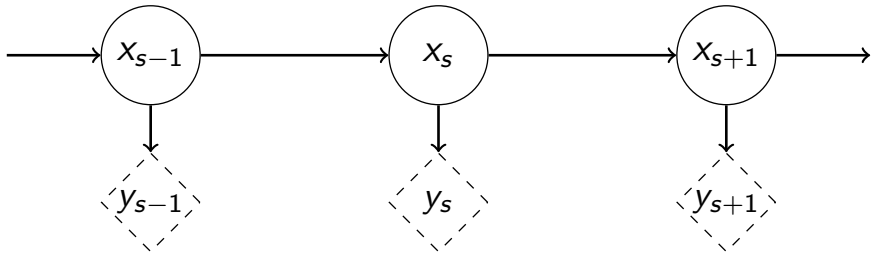
Bayesian Networks: popularity

Hidden Markov Models (Baum et al. 1966) (Dymarski 2011)
Bayesian Networks with latent $\boldsymbol{X}$ and observed variables $\boldsymbol{Y}$. Widely known for:

- Stock index forecasting (Gorynin et al. 2017)
- Speech processing (Toda 2011)
- Gene prediction (Stanke et al. 2006)

Hidden Markov Chain with Independent Noise:

$$p(\boldsymbol{x}, \boldsymbol{y}) = p(x_r) \prod_{s \in \bar{\mathcal{S}}} p(x_s | x_{s-1}) \prod_{s \in \mathcal{S}} p(y_s | x_s) \tag{2}$$



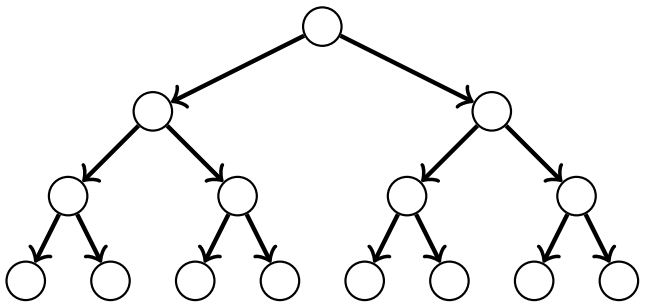
Hidden Markov Chain ***(direct computations)***

The probabilistic models

Markov Tree (MT) Bayesian Networks

Vertices (except roots) have one parent.
MT joint distribution (Monfrini et al. 2003):

$$p(\boldsymbol{x}) = p(x_r) \prod_{s \in \bar{\mathcal{S}}} p(x_s | x_{s^-}). \qquad (3)$$



Dyadic MT (3 layers)
(direct computations)

- *Dyadic* tree: 1 root and 2 sons/vertice (except last layer).
- *Quadtree*: 1 root and 4 sons/vertice (except last layer).

Exact inference dyadic/quad trees:

- Upward-Downward (Durand et al. 2004) (principled approach to get the marginals in dyadic MT).

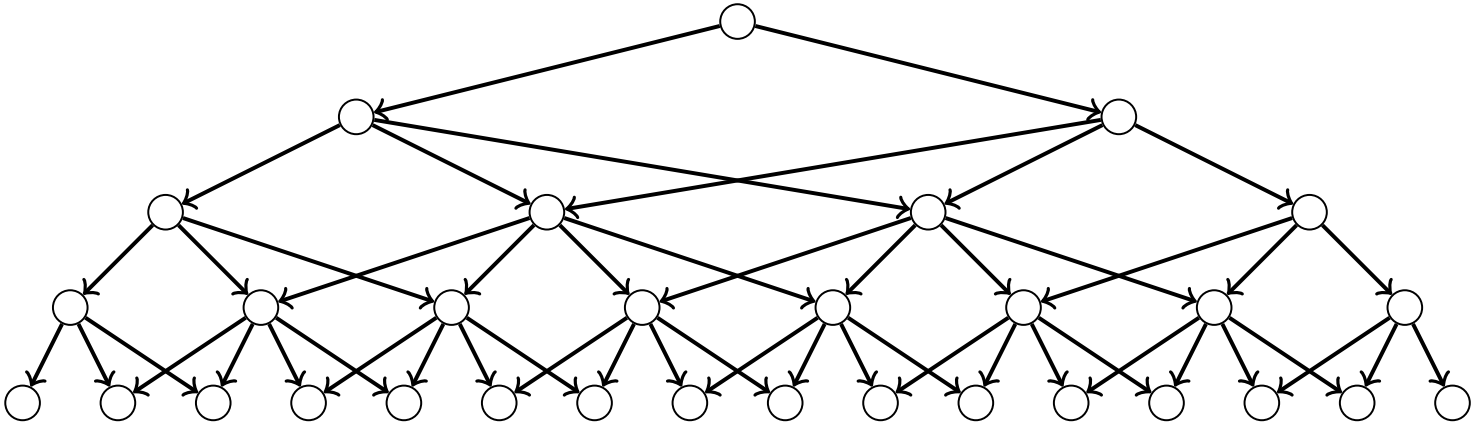
$$p(x_s) = \sum_{x_{s-}} p(x_{s-}) p(x_s | x_{s-}). \qquad (4)$$

Spatial Bayesian Networks (SBN)

A SBN is a Bayesian network with joint distribution **(this paper)**:

$$p(\boldsymbol{x}) = p(x_r) \prod_{s \in \bar{\mathcal{S}}} p(x_s | x_{s^-}, x_{v(s)}), \tag{5}$$

v associates a father's neighbour: not a Markov Tree!
The *spatial context* is modeled but the numerous loops makes inference complex and approximate.



SBN (based on a dyadic tree) with 4 layers
(approximate computations)

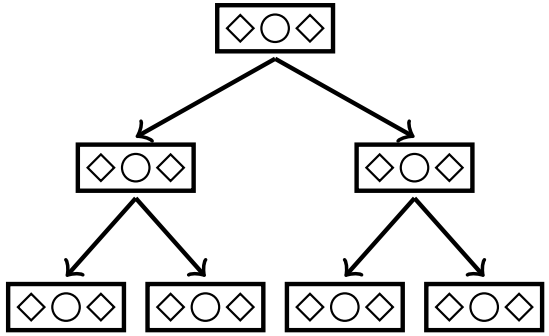
Spatial Triplet Markov Tree Bayesian Networks

In STMT, $\boldsymbol{T} = (\boldsymbol{X}, \boldsymbol{V})$ and $\boldsymbol{V}$ auxiliary process **(this paper)** and (Courbot et al. 2018).

$\boldsymbol{T}$ is a *Markov Tree* with joint distribution:

$$p(\boldsymbol{t}) = p(\boldsymbol{t}_r) \prod_{s \in \bar{\mathcal{S}}} p(\boldsymbol{t}_s | \boldsymbol{t}_{s-}), \tag{6}$$

with $\boldsymbol{t_s}$ a triplet $(x_s, v_s^{\leftarrow}, v_s^{\rightarrow})$



STMT (Markov property)

(direct computations)

Spatial Triplet Markov Tree Bayesian Networks

In STMT, $\boldsymbol{T} = (\boldsymbol{X}, \boldsymbol{V})$ and $\boldsymbol{V}$ auxiliary process **(this paper)** and (Courbot et al. 2018).

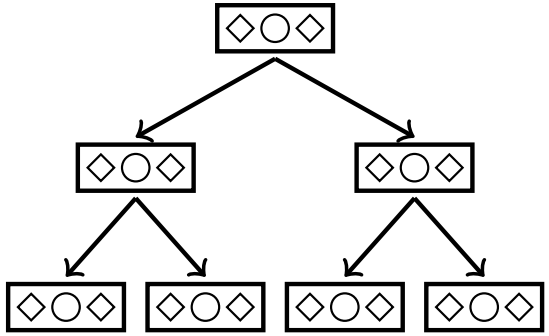
$\boldsymbol{T}$ is a *Markov Tree* with joint distribution:

$$p(\boldsymbol{t}) = p(\boldsymbol{t}_r) \prod_{s \in \bar{\mathcal{S}}} p(\boldsymbol{t}_s | \boldsymbol{t}_{s-}), \tag{6}$$

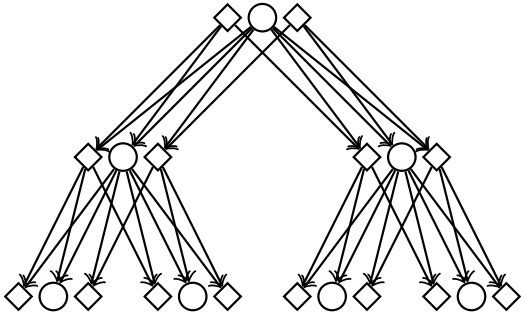
with $\boldsymbol{t_s}$ a triplet $(x_s, v_s^{\leftarrow}, v_s^{\rightarrow})$ such that:

$$p(\boldsymbol{t}_s | \boldsymbol{t}_{s-}) = p(x_s | x_{s-}, \boldsymbol{v}_{s-}) p(v_s^{\leftarrow} | x_{s-}, \boldsymbol{v}_{s-}) p(v_s^{\rightarrow} | x_{s-}, \boldsymbol{v}_{s-}). \tag{7}$$

The transitions are specially designed to model the *spatial context*.



STMT (Markov property)
(direct computations)



STMT (all edges)
(direct computations)

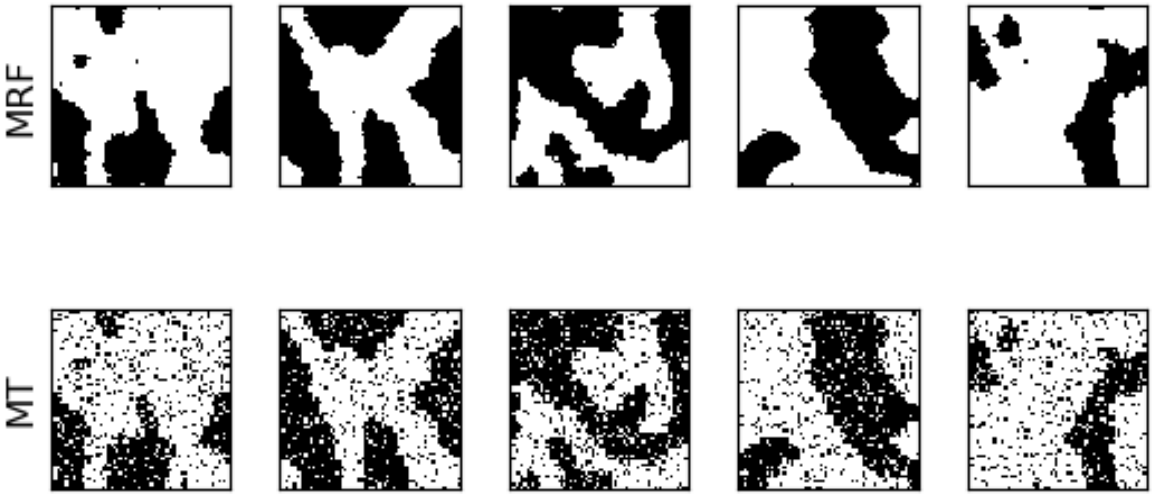
Sampling from the models

Test the ability to capture spatial context by *clamped* sampling from Markov Random Field (MRF) realizations

Average **different rate from MRF**

Sampling from the models

Test the ability to capture spatial context by *clamped* sampling from Markov Random Field (MRF) realizations

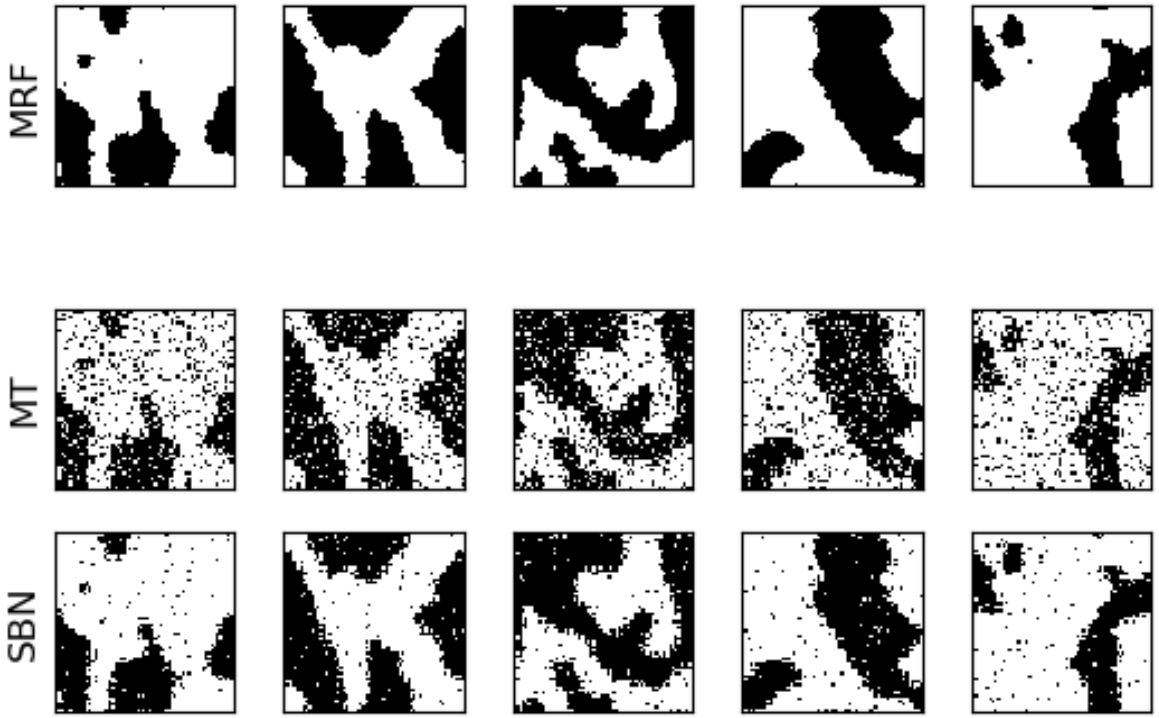


Average **different rate from MRF**

$\rightarrow 0.142$

Sampling from the models

Test the ability to capture spatial context by *clamped* sampling from Markov Random Field (MRF) realizations



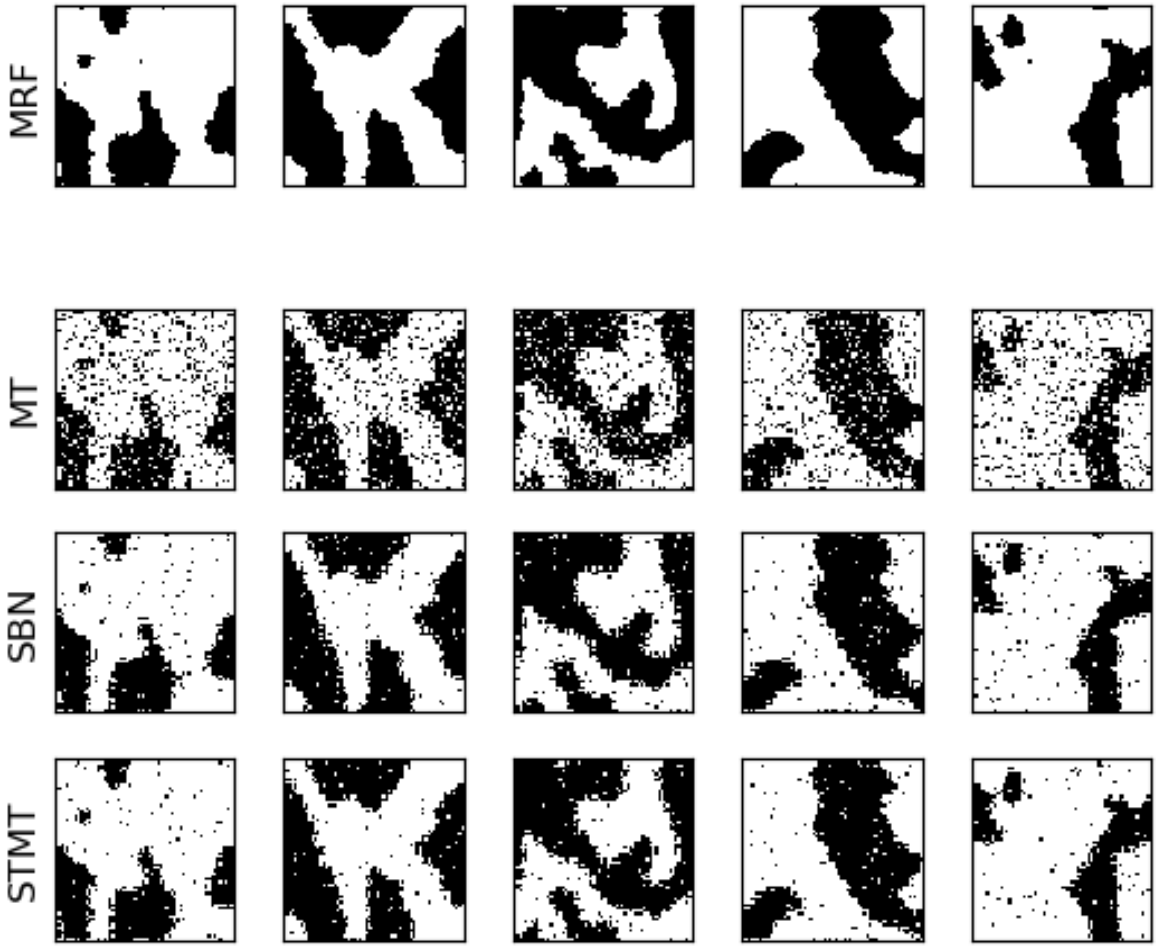
Average **different rate from MRF**

$\rightarrow 0.142$

$\rightarrow 0.072$

Sampling from the models

Test the ability to capture spatial context by *clamped* sampling from Markov Random Field (MRF) realizations



Average **different rate from MRF**

$\rightarrow 0.142$

$\rightarrow 0.072$

$\rightarrow 0.068$

- SBN and STMT seem to capture the *spatial context*

Variational Inference

The method

VI recasts inference into a maximization problem (Jordan et al. 1999):

$$-\mathbb{KL}(q(\boldsymbol{x})||p(\boldsymbol{x})) = \mathbb{E}_q[\log p(\boldsymbol{x})] - \mathbb{E}_q[\log q(\boldsymbol{x})], \tag{8}$$

where:

- p is the *target* distribution (SBN here)
- q is the *variational* distribution (**structured** or not)

The method

VI recasts inference into a maximization problem (Jordan et al. 1999):

$$-\mathbb{KL}(q(\boldsymbol{x})||p(\boldsymbol{x})) = \mathbb{E}_q[\log p(\boldsymbol{x})] - \mathbb{E}_q[\log q(\boldsymbol{x})], \tag{8}$$

where:

- p is the *target* distribution (SBN here)
- q is the *variational* distribution (**structured** or not)

Extended to integrate auxiliary variables (Agakov et al. 2004)

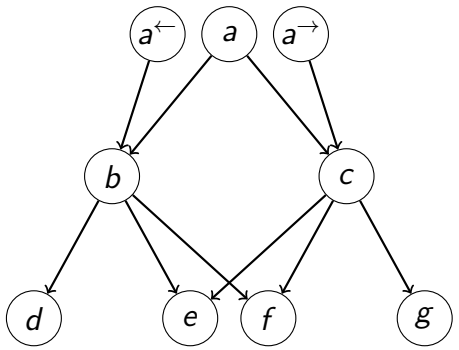
$$\begin{aligned} -\mathbb{KL}(q(\boldsymbol{x},\boldsymbol{v})||\tilde{p}(\boldsymbol{x},\boldsymbol{v})) &= \mathbb{E}_q[\log \tilde{p}(\boldsymbol{x},\boldsymbol{v})] - \mathbb{E}_q[\log q(\boldsymbol{x},\boldsymbol{v})], \\ &= \mathbb{E}_q[\log \tilde{p}(\boldsymbol{x})] + \mathbb{E}_q[\log \tilde{p}(\boldsymbol{v}|\boldsymbol{x})] - \mathbb{E}_q[\log q(\boldsymbol{x},\boldsymbol{v})]. \end{aligned} \tag{9}$$

$\tilde{p}$ is extended with auxiliary variables (with condition: $\tilde{p}(\boldsymbol{x}) = p(\boldsymbol{x})$)

A small SBN example I

- Target distribution:

Let $\boldsymbol{x} = (a, a^{\leftarrow}, a^{\rightarrow}, b, c, d, e, f, g)$,
$p(\boldsymbol{x})$ has a SBN factorization

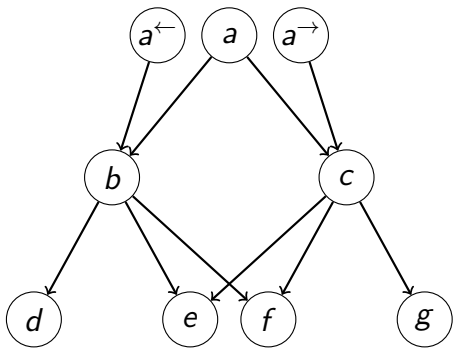


(approximate computations)

A small SBN example I

- Target distribution:

Let $\boldsymbol{x} = (a, a^{\leftarrow}, a^{\rightarrow}, b, c, d, e, f, g)$,
$p(\boldsymbol{x})$ has a SBN factorization



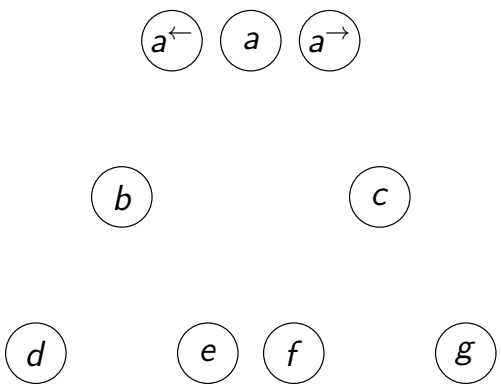
(approximate computations)

- Variational approximation:

Mean-Field (MF) approximation:

$$q^{MF}(\boldsymbol{x}) = \prod_i q(x_i) \qquad (10)$$

$\rightarrow$ Independent random variables

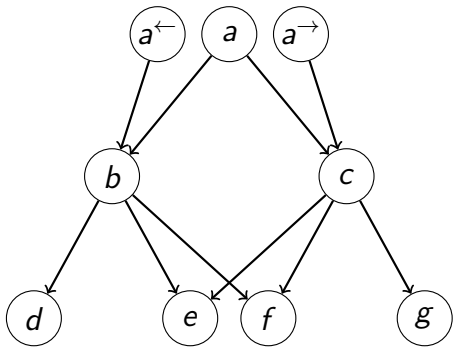


MF approximation
(direct computations)

A small SBN example I

- Target distribution:

Let $\boldsymbol{x} = (a, a^{\leftarrow}, a^{\rightarrow}, b, c, d, e, f, g)$,
$p(\boldsymbol{x})$ has a SBN factorization



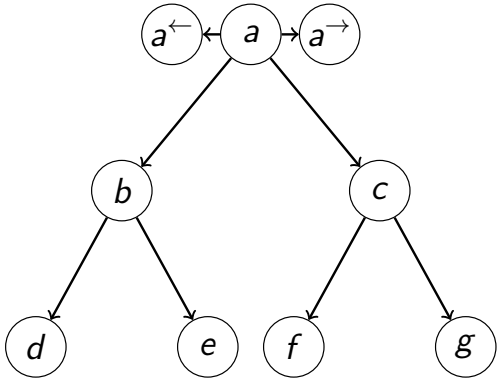
(approximate computations)

- Variational approximation:

Markov Tree (MT) approximation:

$$q^{MT}(\boldsymbol{x}) = q(x_r) \prod_{s \in \bar{\mathcal{S}}} q(x_s | x_{s^-}). \quad (10)$$

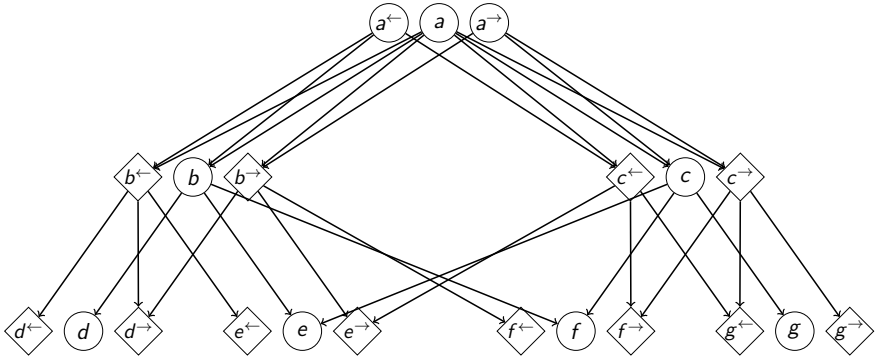
$\rightarrow$ q^{MT} is MT structured



MT approximation
(direct computations)

A small SBN example II

- Target distribution:



(approximate computations)

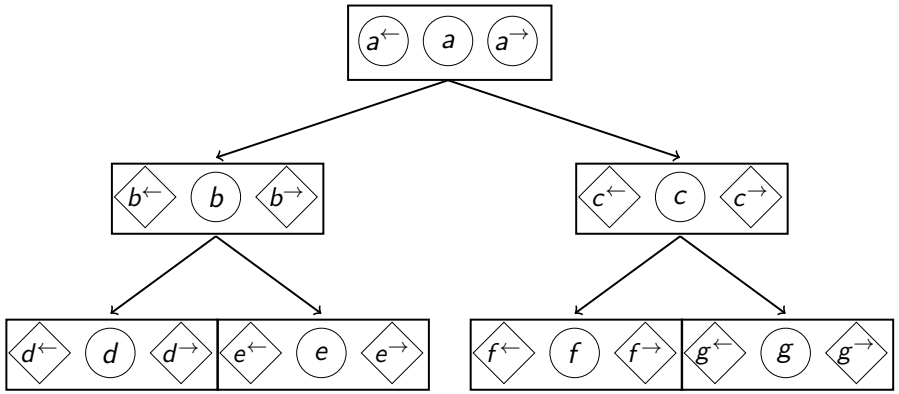
$\tilde{p}(\boldsymbol{x}, \boldsymbol{v})$ is a SBN with auxiliary variables (and $\tilde{p}(\boldsymbol{x}) = p(\boldsymbol{x})$)

- Variational approximation:

STMT approximation:

$$q^{STMT}(\boldsymbol{t}) = q(\boldsymbol{t}_r) \prod_{s \in \bar{\mathcal{S}}} q(\boldsymbol{t}_s | \boldsymbol{t}_{s-}) \tag{11}$$

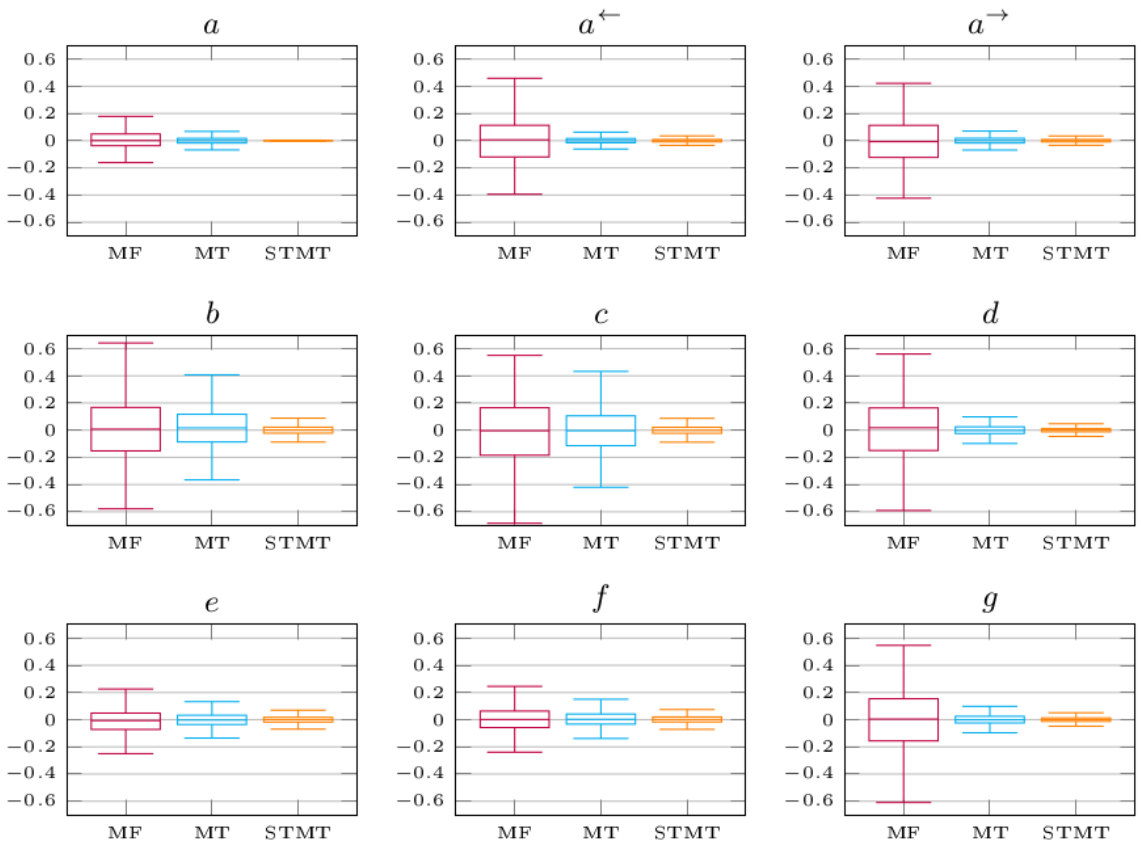
$\rightarrow$ q^{STMT} is STMT structured

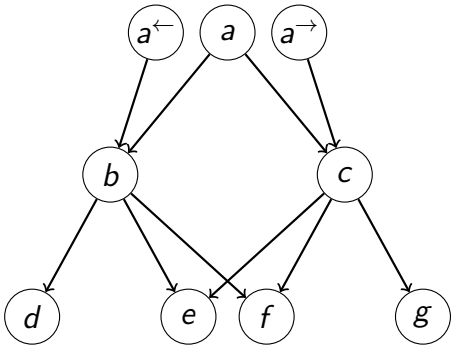


STMT approximation

(direct computations)

Small network example: model proximity

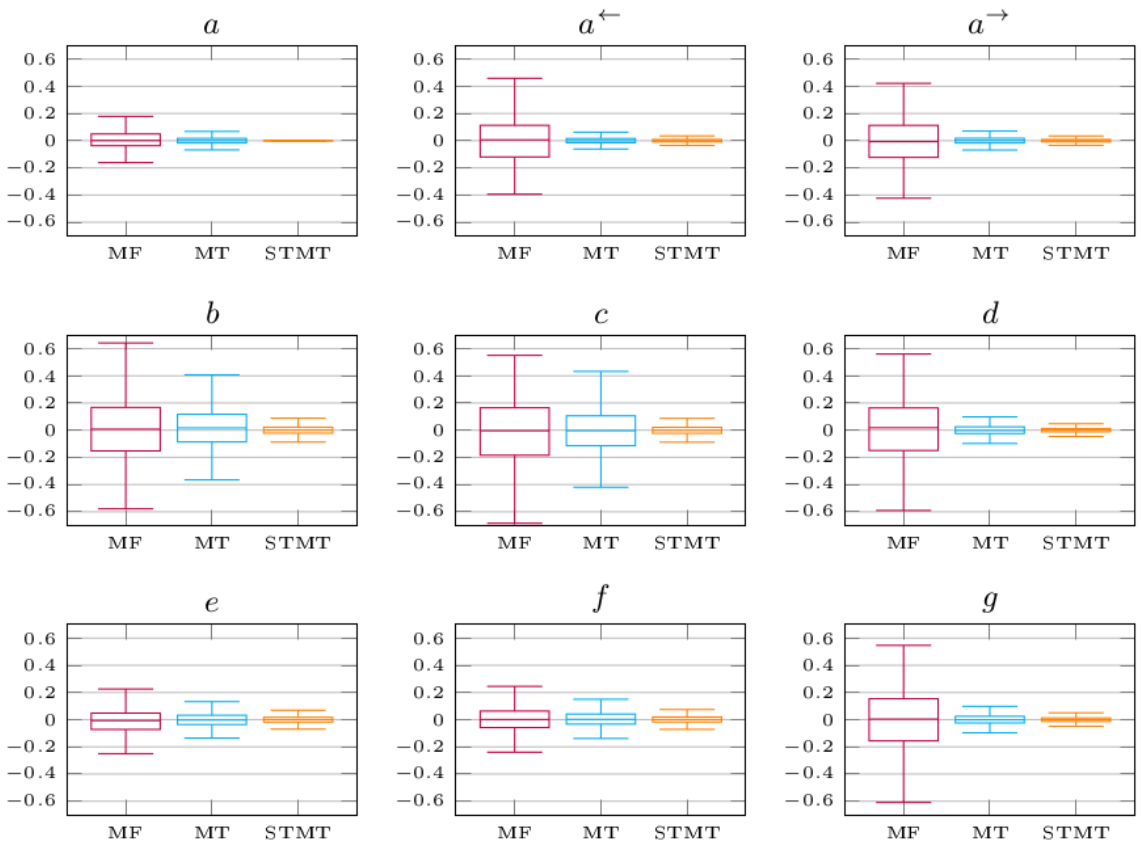


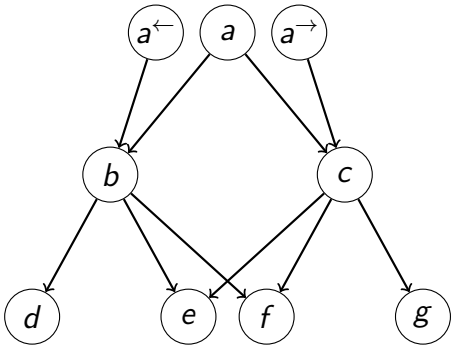


SBN

→ Dispersions of errors for true marginals estimation (1000 trials)

Small network example: model proximity





SBN

→ Dispersions of errors for true marginals estimation (1000 trials)

- Error dispersion : MF VI > MT VI > STMT VI
- STMT seems to best capture the enhanced correlations of SBN

Conclusion & Perspectives

Conclusion & Perspectives

Conclusions:

- 2 new tree-structured models with rich correlations
- Importance of the triplet structure for exact approaches in more complex networks.
- Auxiliary variables in probabilistic models can be an efficient asset if the model is well-thought!

Perspectives:

- Analytical comparison STMT/SBN
- Extend the work to quadtrees for image processing

References I

[1] Felix V Agakov and David Barber. "An auxiliary variational method". In: *International Conference on Neural Information Processing*. Springer. 2004, pp. 561–566.

[2] Leonard E Baum and Ted Petrie. "Statistical inference for probabilistic functions of finite state Markov chains". In: *The annals of mathematical statistics* 37.6 (1966), pp. 1554–1563.

[3] Jean-Baptiste Courbot et al. "Triplet markov trees for image segmentation". In: *SSP 2018: IEEE Workshop on Statistical Signal Processing*. 2018, pp. 233–237.

[4] J-B Durand, Paulo Goncalves, and Yann Guédon. "Computational methods for hidden Markov tree models-An application to wavelet trees". In: *IEEE Transactions on Signal Processing* 52.9 (2004), pp. 2551–2560.

[5] Przemyslaw Dymarski. *Hidden Markov Models: Theory and Applications*. IntechOpen, 2011.

[6] Stuart Geman and Donald Geman. "Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images". In: *IEEE Transactions on pattern analysis and machine intelligence* 6 (1984), pp. 721–741.

[7] Ivan Gorynin, Emmanuel Monfrini, and Wojciech Pieczynski. "Pairwise Markov models for stock index forecasting". In: *25th European Signal Processing Conference (EUSIPCO)*. IEEE. 2017, pp. 2041–2045.

[8] Michael I Jordan et al. "An introduction to variational methods for graphical models". In: *Machine learning* 37.2 (1999), pp. 183–233.

[9] E Monfrini et al. "Image and signal restoration using pairwise Markov trees". In: *IEEE Workshop on Statistical Signal Processing, 2003*. IEEE. 2003, pp. 174–177.

[10] Judea Pearl. "Reverend bayes on inference engines: a distributed hierarchical approach". In: *Proceedings of the Second AAAI Conference on Artificial Intelligence*. 1982, pp. 133–136.

References II

[11] Mario Stanke et al. "Gene prediction in eukaryotes with a generalized hidden Markov model that uses hints from external sources". In: *BMC bioinformatics* 7.1 (2006), p. 62.

[12] Ruslan Leont'evich Stratonovich. "Conditional markov processes". In: *Non-linear transformations of stochastic processes*. Elsevier, 1965, pp. 427–453.

[13] Tomoki Toda. "Modeling of speech parameter sequence considering global variance for HMM-based speech synthesis". In: *Hidden Markov Models, Theory and Applications*. InTech, 2011, pp. 131–150.

[14] Yair Weiss. "Correctness of local probability propagation in graphical models with loops". In: *Neural computation* 12.1 (2000), pp. 1–41.