Bayesian Learning on Graphs using Deep Gaussian Markov Random Fields

NORDSTAT 2023: *Gaussian Processes on Networks and Graphs*

Joel Oskarsson, Per Sidén, Fredrik Lindsten

Division of Statistics and Machine Learning
Department of Computer and Information Science
Linköping University, Sweden

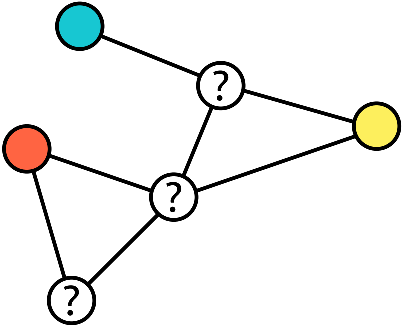
LINKÖPING UNIVERSITY

Based on papers

Joel Oskarsson, Per Sidén, and Fredrik Lindsten. "**Scalable Deep Gaussian Markov Random Fields for General Graphs**". In: *Proceedings of the 39th International Conference on Machine Learning*. PMLR. 2022

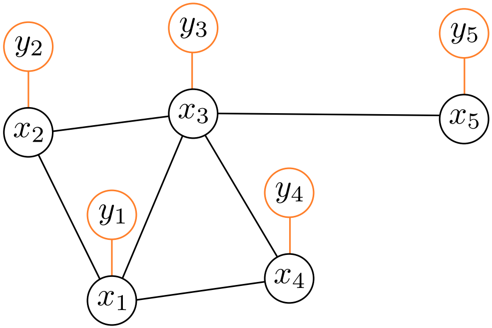
Per Sidén and Fredrik Lindsten. "**Deep Gaussian Markov Random Fields**". In: *Proceedings of the 37th International Conference on Machine Learning*. PMLR. 2020

Our main setting



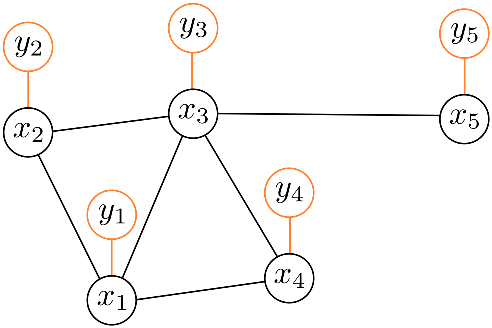
- Node prediction
- Single, fully known graph structure
- Subset of nodes observed

Probabilistic model



$$
\begin{aligned}
p(\boldsymbol{x}, \boldsymbol{y}) &= p(\boldsymbol{x}) p(\boldsymbol{y}|\boldsymbol{x}) \\
&= p(\boldsymbol{x}) \prod_{i=1}^{N} p(y_i|x_i)
\end{aligned}
$$

Probabilistic model

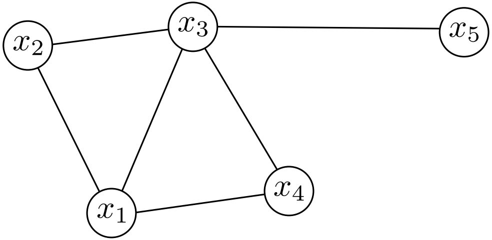


$$p(\boldsymbol{x}, \boldsymbol{y}) = p(\boldsymbol{x})p(\boldsymbol{y}|\boldsymbol{x})$$
$$= p(\boldsymbol{x}) \prod_{i=1}^{N} p(y_i|x_i)$$

- Gaussian prior: $p(\boldsymbol{x}) = \mathcal{N}\left(\boldsymbol{x} \middle| \boldsymbol{\mu}, Q^{-1}\right)$
- Likelihood: $p(y_i|x_i) = \mathcal{N}\left(y_i \middle| x_i, \sigma^2\right)$

- Posterior of interest: $p(\boldsymbol{x}|\boldsymbol{y}_{\text{Obs.}})$

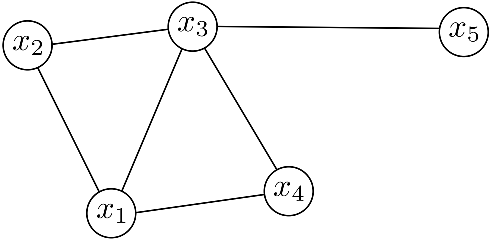
Prior for $\boldsymbol{x}$: Gaussian Markov Random Field (GMRF)[1]



- Markov random fields
  - Undirected graphical models

[1] Håvard Rue and Leonhard Held. *Gaussian Markov random fields: theory and applications*. Monographs on statistics and applied probability 104. Chapman & Hall/CRC, 2005

Prior for $\boldsymbol{x}$: Gaussian Markov Random Field (GMRF)[1]



- Markov random fields
  - Undirected graphical models
- GMRFs: joint Gaussian density
  - $p(\boldsymbol{x}) = \mathcal{N}\left(\boldsymbol{x} \middle| \boldsymbol{\mu}, Q^{-1}\right)$
  - Precision matrix Q
  - Non-zeros in Q = edges in graph
  - Can have sparse Q but dense Q^{-1}

[1] Håvard Rue and Leonhard Held. *Gaussian Markov random fields: theory and applications*. Monographs on statistics and applied probability 104. Chapman & Hall/CRC, 2005

Deep Gaussian Markov Random Field (DGMRF)[1] prior

- Define GMRF prior using affine map g

$$g(\boldsymbol{x}) = G\boldsymbol{x} + \boldsymbol{b} = \boldsymbol{z}, \quad \boldsymbol{z} \sim \mathcal{N}(\boldsymbol{0}, I) \qquad \Rightarrow \qquad \boldsymbol{x} \sim \mathcal{N}\left(-G^{-1}\boldsymbol{b}, (G^{\intercal}G)^{-1}\right)$$

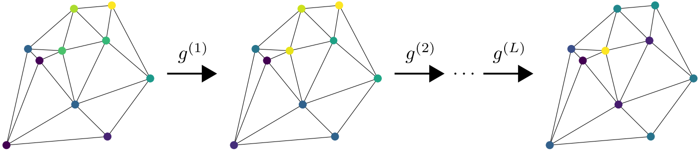
[1] Per Sidén and Fredrik Lindsten. "Deep Gaussian Markov Random Fields". In: *Proceedings of the 37th International Conference on Machine Learning*. PMLR. 2020

Deep Gaussian Markov Random Field (DGMRF)[1] prior

- Define GMRF prior using affine map g

$$g(\boldsymbol{x}) = G\boldsymbol{x} + \boldsymbol{b} = \boldsymbol{z}, \quad \boldsymbol{z} \sim \mathcal{N}(\mathbf{0}, I) \qquad \Rightarrow \qquad \boldsymbol{x} \sim \mathcal{N}\left(-G^{-1}\boldsymbol{b}, (G^{\intercal}G)^{-1}\right)$$

- Layered structure $g(\boldsymbol{x}) = g^{(L)} \circ g^{(L-1)} \circ \cdots \circ g^{(1)}(\boldsymbol{x})$

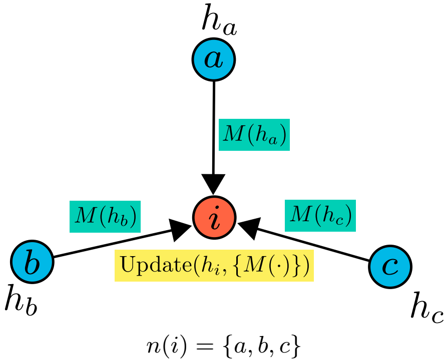


$\boldsymbol{x} \sim \mathcal{N}\left(\boldsymbol{\mu}, Q^{-1}\right)$ $\qquad$ $\boldsymbol{z} \sim \mathcal{N}(\mathbf{0}, I)$

[1] Per Sidén and Fredrik Lindsten. “Deep Gaussian Markov Random Fields”. In: *Proceedings of the 37th International Conference on Machine Learning*. PMLR. 2020

A (very brief) introduction to Graph Neural Networks (GNNs)

- Message passing neural networks[1]
  - Node representation h
  - Send messages
  - Update nodes

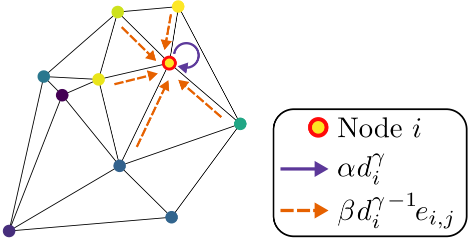


[1] Justin Gilmer et al. "Neural Message Passing for Quantum Chemistry". In: *Proceedings of the 34th International Conference on Machine Learning*. PMLR. 2017

A DGMRF layer for graphs[1]

$$h_i^{(l+1)} = g^{(l+1)}\left(\boldsymbol{h}^{(l)}\right)_i$$
$$= b + \alpha d_i^{\gamma} h_i^{(l)} + \sum_{j \in n(i)} \beta d_i^{\gamma-1} e_{j,i} h_j^{(l)}$$

- Trainable parameters $\theta = \{\alpha, \beta, \gamma, b\}$
- Implemented as GNN



[1] Joel Oskarsson, Per Sidén, and Fredrik Lindsten. "Scalable Deep Gaussian Markov Random Fields for General Graphs". In: *Proceedings of the 39th International Conference on Machine Learning*. PMLR. 2022

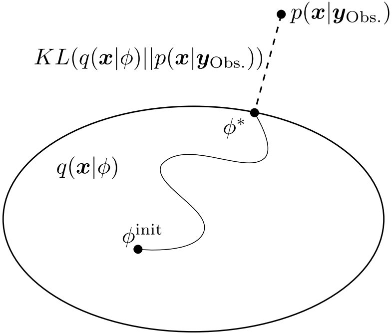
Learning parameters of the DGMRF prior

- Empirical Bayes
- Maximize log marginal likelihood, $\log p(\boldsymbol{y}_{\text{Obs.}}|\theta)$
  - Iterative, gradient-based optimization
  - *Infeasible!*

Learning parameters of the DGMRF prior

- Empirical Bayes
- Maximize log marginal likelihood, $\log p(\boldsymbol{y}_{\mathrm{Obs.}}|\theta)$
  - Iterative, gradient-based optimization
  - *Infeasible!*

- Variational inference to the rescue!
  - Maximize the Evidence Lower Bound (ELBO)
  - Variational distribution: $q(\boldsymbol{x}|\phi)$



Learning parameters of the DGMRF prior (cont.)

- With learned parameters:
  - Exact posterior inference **once**

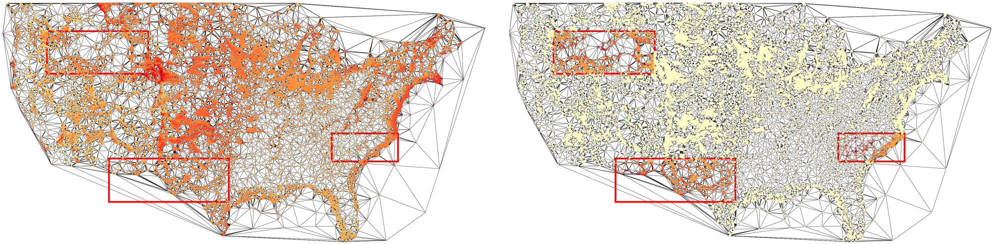
$$p(\boldsymbol{x}|\boldsymbol{y}_{\text{Obs.}}) = \mathcal{N}\left(\boldsymbol{x}\middle|\tilde{\boldsymbol{\mu}}, \tilde{Q}^{-1}\right) \qquad \tilde{\boldsymbol{\mu}} = \tilde{Q}^{-1}\left(Q\boldsymbol{\mu} + \frac{1}{\sigma^2}\boldsymbol{y}_{\text{Obs.}}\right) \qquad \tilde{Q} = Q + \frac{1}{\sigma^2} I_{\text{Obs.}}$$

  - Conjugate Gradient method to solve $\tilde{Q}^{-1}$ and draw posterior samples[1]

[1] George Papandreou and Alan L Yuille. "Gaussian sampling by local perturbations". In: *Advances in Neural Information Processing Systems* 23 (2010)

Experiment: Wind speed data

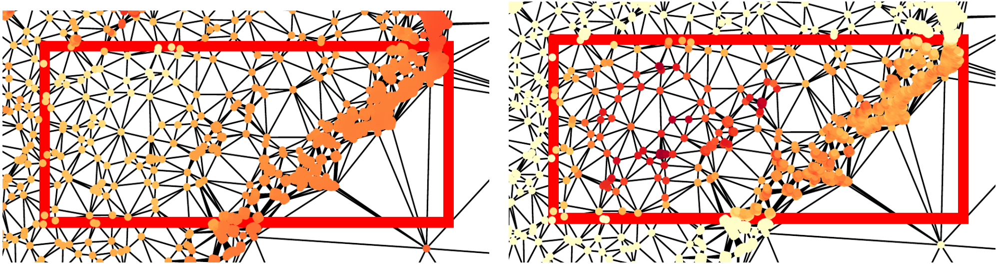
- Wind speeds at 130 000 sites in the US
- Weighted spatial graph based on coordinates
- Nodes in red rectangles are unobserved



(a) Posterior mean (b) Posterior marginal std.-dev.

Experiment: Wind speed data

- Wind speeds at 130 000 sites in the US
- Weighted spatial graph based on coordinates
- Nodes in red rectangles are unobserved



(a) Posterior mean (b) Posterior marginal std.-dev.

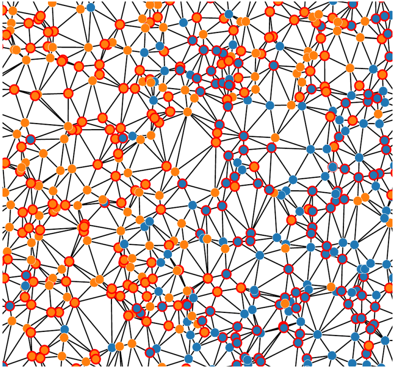
Experiment: Wikipedia graphs

- Nodes = 11 000 Wikipedia pages[1]
  - 50% observed
- Edges = links between pages
- $\boldsymbol{y}$ = log average monthly traffic

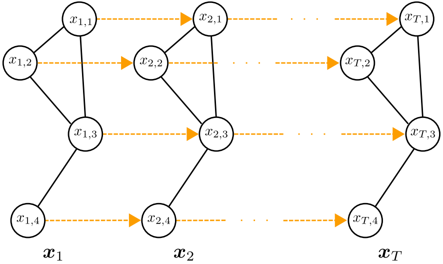
	RMSE	–CRPS
Baselines		
Matérn Graph GP	2.169	1.251
DGP (GNN ensemble)	1.308	0.786
I-GMRF	1.526	0.939
DGMRF		
1 Layer	1.311	0.704
3 Layer	1.228	0.652
5 Layer	**1.169**	**0.614**

[1] Benedek Rozemberczki, Carl Allen, and Rik Sarkar. "Multi-scale attributed node embedding". In: *Journal of Complex Networks* 9.2 (2021)

Ongoing and future work



Non-Gaussian likelihoods: node classification



Spatio-temporal data

Summary

- Deep Gaussian Markov Random Field (DGMRF) models on graphs
- DGMRF layers based on graph neural networks
- Efficient and scalable parameter learning and inference
- Code available: `github.com/joeloskarsson/graph-dgmrf`

Joel Oskarsson

joel.oskarsson@liu.se

joeloskarsson.github.io

Per Sidén

psiden@qti.qualcomm.com

Fredrik Lindsten

fredrik.lindsten@liu.se

lindsten.netlify.app