Graph kernels via SPDE framework

Alexander Nikitin, Aalto University

Preliminaries

- Graph: a set of nodes and a set of edges,
- Examples: molecules, telecom networks, ...,
- Spatial and spatio-temporal problems on graphs,
- Uncertainty quantification is crucial for many applications,
- In this presentation the graph is static and the signal changes.

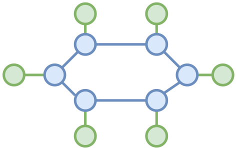


Figure: Molecule graph C_6H_6

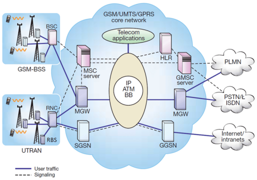
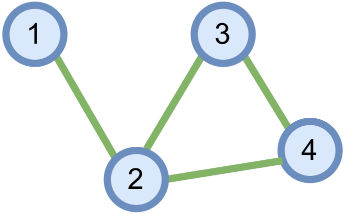
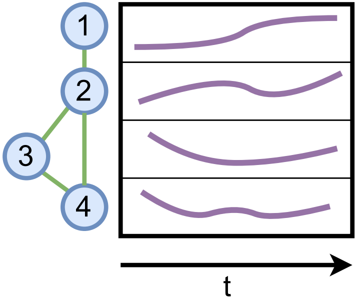


Figure: Cable Network[1]

[1] Lars-Örjan Kling (2002). ”CPP—Cello packet platform”. Ericsson Review (Ericsson AB) (2)

Problem





Goal: develop spatial and spatio-temporal probabilistic methods on graphs.

Preliminaries: Graphs

- Graph: $G = (V, E)$,
- Adjacency matrix $\boldsymbol{A}$:

$$\boldsymbol{A}_{(i,j)} = 1$$

- Weight matrix $\boldsymbol{W}$:

$$\boldsymbol{W}_{(i,j)} = \text{weight(i, j)}$$

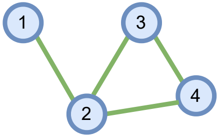
- Degree:

$$\boldsymbol{D}_W = \text{diag}(w_i)$$

- The graph Laplacian $\boldsymbol{L}$:

$$\boldsymbol{L} = \boldsymbol{D}_W - \boldsymbol{W},$$

where $w_i = \sum_{j:(i,j)\in E} w_{ij}$



A

0	1	0	0
1	0	1	1
0	1	0	1
0	1	1	0

D

1	0	0	0
0	3	0	0
0	0	2	0
0	0	0	2

Figure: Graph, adjacency matrix and degree matrix

Gaussian Processes

- Motivation: non-linear regression,
- Prior over functions and likelihood:
  $p(f|\theta) = \mathcal{GP}\left(f; 0,\ \mathrm{K}_\theta\right); p\left(y_n|f, x_n, \theta\right)$
- Learning: infer underlying function given the dataset and estimate marginal likelihood
  $p(f|\mathbf{y}, \mathbf{x}, \theta), p(\mathbf{y}|\mathbf{x}, \theta)$,
- Pros: data efficiency, tractability, probabilistic interpretability.
- Cons: computational complexity, empirically often worse than neural networks (for high-dimensional problems).
- In order to apply to new domains, develop a kernels.

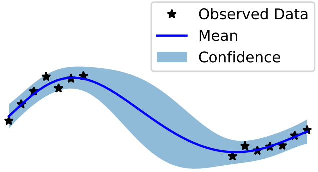


Figure: Gaussian process

Question: How to develop kernels for graph nodes?

Graph Kernels

- Kernel over a finite space,
- Any symmetric semi-definite matrix works,
- Can we incorporate extra knowledge in a kernel?
  - topology,
  - locality.
- Possible approaches:
  - Geometric,
  - Functional, $||f||_K$ should be small when f is smooth,
  - Discretization of continuous kernel.
- Embedding distances does not work for arbitrary graphs:

$$\frac{1}{(4\pi t)^{\frac{d}{2}}} \exp\left(-\frac{\|\mathbf{x}-\mathbf{x}_0\|^2}{4t}\right) \tag{1}$$

Graph Kernels: Diffusion

- Let us consider an equation

$$\frac{\partial}{\partial t} K(\mathbf{x}, t) = \Delta K(\mathbf{x}, t) \tag{2}$$

Subject to initial condition

$$K(\boldsymbol{x}, 0) = \delta(\boldsymbol{x}). \tag{3}$$

.

- Solution (normalized Gaussian kernel):

$$K(\mathbf{x}, t) = K(\mathbf{x}_0, \mathbf{x}) = \frac{1}{(4\pi t)^{\frac{d}{2}}} \exp\left(-\frac{\|\mathbf{x} - \mathbf{x}_0\|^2}{4t}\right) \tag{4}$$

- Similarly the equation can be considered on graphs (Smola and Kondor, 2003)

$$\frac{\partial}{\partial t} \boldsymbol{f}_t = -\boldsymbol{L}\boldsymbol{f}_t, \tag{5}$$

$$\boldsymbol{f}_t = \boldsymbol{f}_0 e^{-t\boldsymbol{L}} \tag{6}$$

From Graph Kernels to spatio-temporal kernels

- Spatio-temporal prediction $f : V \times \mathbb{R} \to \mathbb{R}$
- Product of spatial and temporal kernel: $K(x, x', t, t') = K(x, x') \times K(t, t')$,
- Product-separable kernels are not as expressive as non-separable,

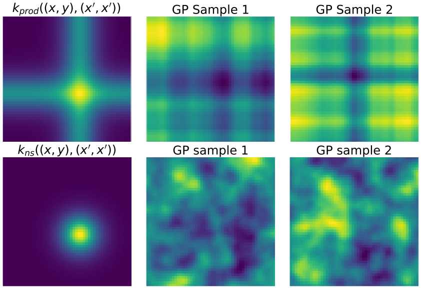


Figure: Prod.-separable and non-separable Matérn 5/2

Can we do better? SPDE approach.

- SPDE Matérn kernels (Lindgren, Rue, and Lindström, 2011). Gaussian field $x(\boldsymbol{u})$ with Matérn kernels is a solution to the linear fractional SPDE:

$$\left(\kappa^2 - \Delta\right)^{\alpha/2} x(\mathbf{u}) = \mathcal{W}(\mathbf{u}),$$
$$\mathbf{u} \in \mathbb{R}^d, \alpha = \nu + d/2, \kappa > 0, \nu > 0 \quad (7)$$

- $\Delta \to -\boldsymbol{L}$, gives graph Matérn kernel,
- We apply the same to spatio-temporal SPDEs.

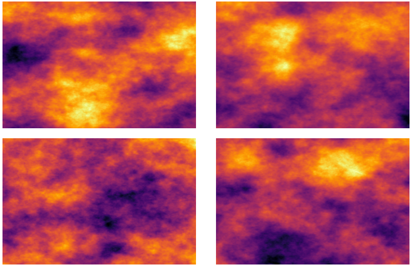
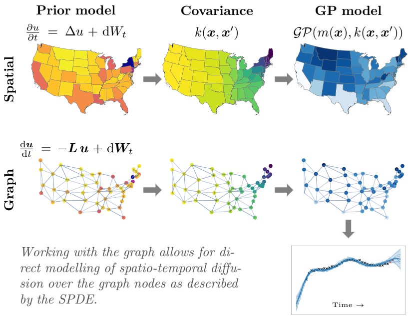


Figure: Two-dimensional Gaussian field with Matérn kernel.

Graph Gaussian processes via SPDEs

(with ST John, Arno Solin, and Samuel Kaski)

- Using the analogues of well-known SPDEs from spatial statistics, we derive non-separable spatio-temporal kernels on graphs
- Empirical evaluation: synthetic datasets, chickenpox and the COVID-19 epidemic forecast



Working with the graph allows for direct modelling of spatio-temporal diffusion over the graph nodes as described by the SPDE.

Framework: SPDE $\rightarrow$ graph kernel

i. Define an SPDE, using prior knowledge about the underlying process
ii. Convert the continuous SPDE to a graph counterpart
iii. Solve the graph counterpart
iv. Derive corresponding mean and covariance function of GP on graph

Instances of the framework

i Matérn kernel

$$\underbrace{\left(\frac{2\nu}{\kappa^2}\boldsymbol{I} + \boldsymbol{L}\right)^{\frac{\nu}{2}}}_{\widetilde{\boldsymbol{L}}} \boldsymbol{u} = \boldsymbol{w}$$

ii Stochastic Heat Equation Kernel (SHEK)

$$\frac{\mathrm{d}\boldsymbol{u}_t}{\mathrm{d}t} = -c\widetilde{\boldsymbol{L}}\boldsymbol{u}_t + \sigma \dot{\boldsymbol{W}}_t$$

iii Stochastic Wave Equation Kernel (SWEK)

$$\frac{\mathrm{d}^2\boldsymbol{u}_t}{\mathrm{d}t^2} = -c^2\widetilde{\boldsymbol{L}}\boldsymbol{u}_t + \sigma \dot{\boldsymbol{W}}_t$$

Results SHEK and SWEK

SHEK

$$\boldsymbol{u}(t) \sim \mathcal{GP}(\boldsymbol{\mu}(t), \mathrm{Cov}[\boldsymbol{u}(s), \boldsymbol{u}(t)]),$$

$$\boldsymbol{\mu}(t) = e^{-c\widetilde{\boldsymbol{L}}t}\boldsymbol{u}(0),$$

$$\mathrm{Cov}[\boldsymbol{u}(t), \boldsymbol{u}(s)] = \frac{\sigma^2}{c} e^{-c\widetilde{\boldsymbol{L}}t - c\widetilde{\boldsymbol{L}}^\top s} \left(e^{c(\widetilde{\boldsymbol{L}}+\widetilde{\boldsymbol{L}}^\top)\min(t,s)} - \boldsymbol{I}\right)(\widetilde{\boldsymbol{L}} + \widetilde{\boldsymbol{L}}^\top)^{-1}.$$

SWEK

$$\boldsymbol{u}(t) \sim \mathcal{GP}(\boldsymbol{\mu}, \mathrm{Cov}[\boldsymbol{u}(s), \boldsymbol{u}(t)],$$

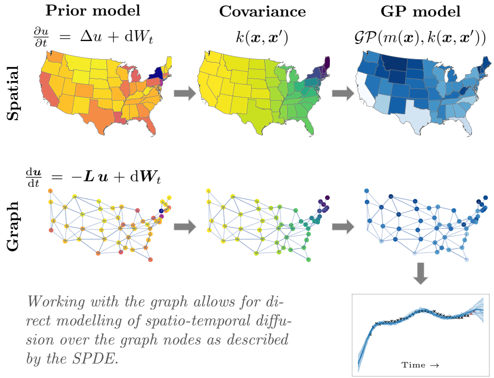
$$\boldsymbol{\mu}(t) = \frac{1}{c}\widetilde{\boldsymbol{L}}^{-\frac{1}{2}} \sin(\boldsymbol{\Theta} t)\boldsymbol{P}\dot{\boldsymbol{u}}(0) + \cos(\boldsymbol{\Theta} t)\boldsymbol{P}\boldsymbol{u}(0),$$

$$\mathrm{Cov}[\boldsymbol{u}(s), \boldsymbol{u}(t)] = \sigma^2\boldsymbol{\Theta}^{-2}\Big(\cos(\boldsymbol{\Theta}(t-s)) \min(t,s) - \frac{1}{2}\cos(\boldsymbol{\Theta}\max(t,s))\sin(\boldsymbol{\Theta}\min(t,s))\boldsymbol{\Theta}^{-1}\Big),$$

where $\boldsymbol{\Theta} = c\sqrt{\widetilde{\boldsymbol{L}}}$ and $\boldsymbol{P}$ is defined using the diagonalization of the fractional Laplacian matrix: $\widetilde{\boldsymbol{L}} = \boldsymbol{P}^{-1}\widetilde{\boldsymbol{L}}_d\boldsymbol{P}$.

COVID-19 graph dataset

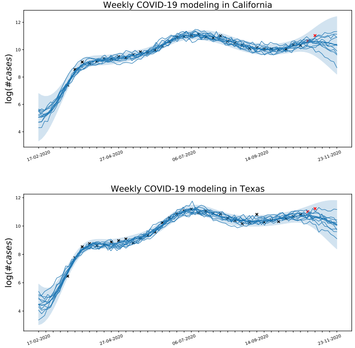
- Each state is a node,
- Connections: adjacencies, flights,
- Target: the number of cases, the number of deaths,
- Normalized and non-normalized by the population,
- Use it: `https://github.com/AlexanderVNikitin/covid19-on-graphs`



Experiments

Tasks (interpolation and extrapolation)

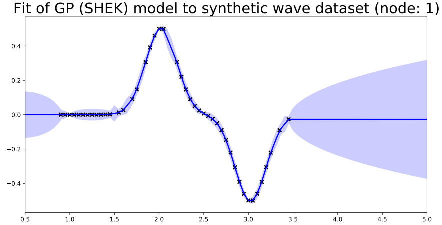
- heat and wave distribution over a one-dimensional line
- spreading of COVID-19 cases across the United States
- spreading of chickenpox cases over Hungarian counties

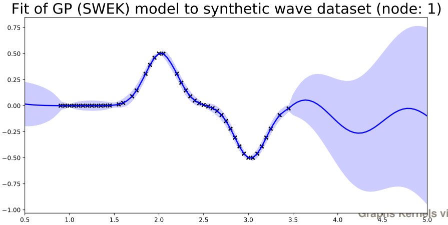


SHEK fit to COVID-19 data.

Experiments: SHEK vs SWEK

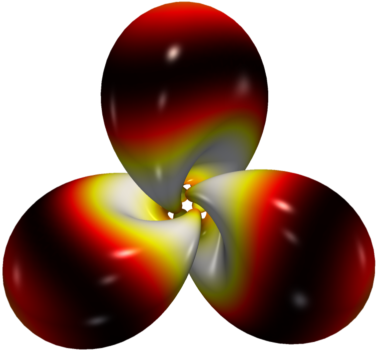
- SHEK: diffusion problems
- SWEK: periodic functions
- Non-separability is an advantage but has computational cost
- the kernel can be chosen using domain knowledge about the problem:
  - Epidemy distribution $\rightarrow$ SHEK,
  - Long-term weather forecasting $\rightarrow$ SWEK





Future directions

- New SPDE types and their application,
- Scaling graph GPs,
- Spatio-temporal models on manifolds,
- Approximation of continuous-domain GPs with graph discretization.



References

- Smola, Alexander J and Risi Kondor (2003). "Kernels and regularization on graphs". In: *Learning Theory and Kernel Machines: 16th Annual Conference on Learning Theory and 7th Kernel Workshop, COLT/Kernel 2003, Washington, DC, USA, August 24-27, 2003. Proceedings*. Springer, pp. 144–158.
- Lindgren, Finn, Håvard Rue, and Johan Lindström (2011). "An explicit link between Gaussian fields and Gaussian Markov random fields: the stochastic partial differential equation approach". In: *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 73.4, pp. 423–498.
- Lindgren, Finn, David Bolin, and Håvard Rue (2022). "The SPDE approach for Gaussian and non-Gaussian fields: 10 years and still running". In: *Spatial Statistics* 50, p. 100599.

Summary

- SPDE framework for kernels on graphs
- SHEK and SWEK kernels for spatio-temporal problems
- Evaluated on synthetic and real (COVID-19 and chickenpox) problems

- This presentation: https://anikitin.me/bayescomp23.pdf
- COVID-19 dataset: https://github.com/AlexanderVNikitin/covid19-on-graphs
- GitHub https://github.com/AaltoPML/spatiotemporal-graph-kernels
- Publication: https://proceedings.mlr.press/v151/nikitin22a.html
- Correspondence: alexander.nikitin@aalto.fi