Coding in Stan: the BYM2 model for disconnected graphs

or

How I learned to stopped worrying and love PC priors

Mitzi Morris

Stan Development Team

Columbia University, New York NY

October 2021

The BYM2 model for disconnected graphs

The BYM2 model is a combination of a GLM plus plus a spatial component which provides localized information pooling between adjacent regions.

- adjacency is a binary relationship
- imposes a graph structure on the set of areal regions.

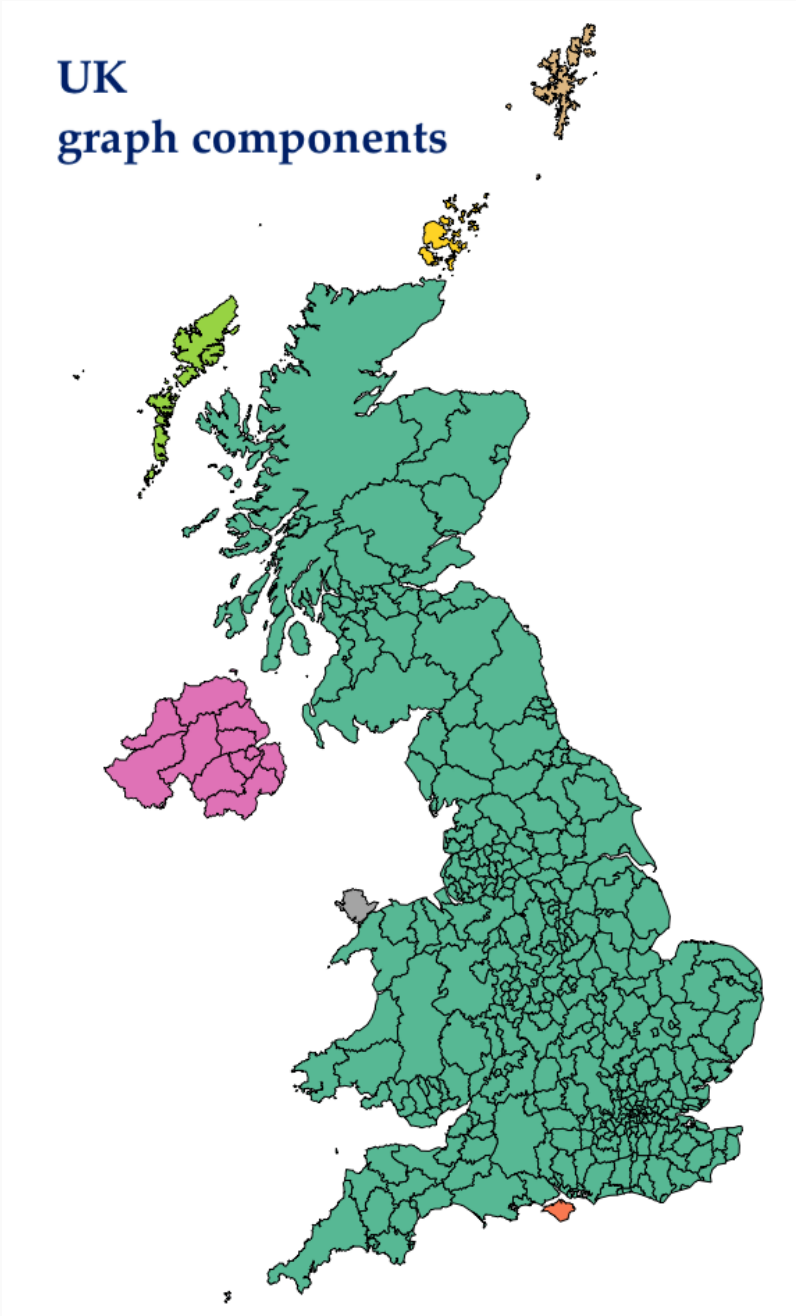
BYM2 model requires a fully connected graph - the analyst must build bridges and tunnels where none exist.

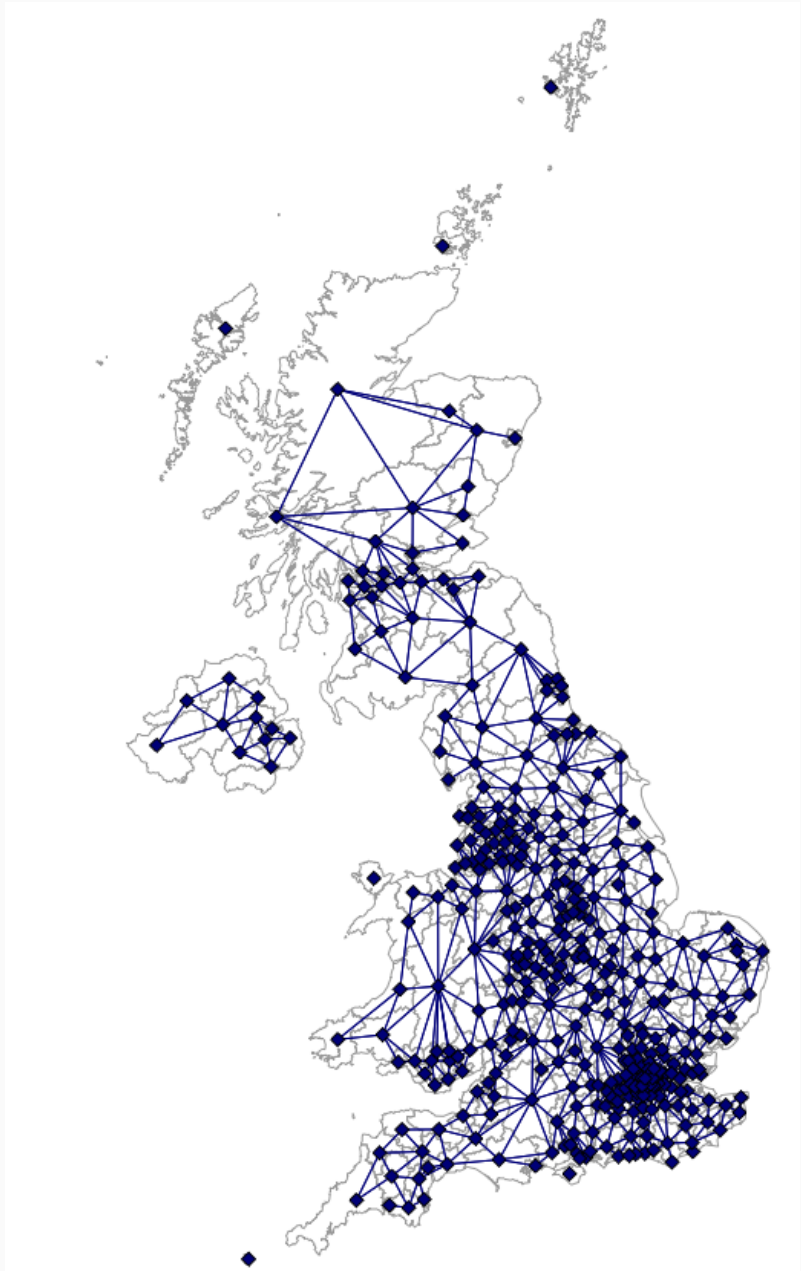
We can extend the model to avoid this.

The Stan language is expressive enough so that the base model can be extended to handle subgraphs and islands.

This talk is about implementation choices and challenges.

Example: United Kingdom = Great Britain + Northern Ireland





GIS Geographic Information Systems - Data and Map Wrangling

To compute adjacency, you need a map (shapefile) and must master one or more GIS software packages.

- Challenge 1: computing the adjacency matrix from the shapefile.
- Challenge 2: computing components and islands for disconnected graphs.
- Challenge 3: aligning areal data structures - node and edge indices with their corresponding outcome and predictor data.

Disease mapping models for areal count data

- Base model: Poisson GLM (interpretable results)
- Poisson plus ICAR (Besag 1973) - accounts for spatial structure
- **BYM** Poisson plus ICAR plus RE (Besag et al 1981) - accounts for spatial and extra-spatial varation
- Subsequent BYM refinements, (Gibbs samplers) Dean et al (), Leroux et al (), Cressie et al ()
- **BYM2** (Riebler et al, 2016) - PC priors for samplers based on joint distribution, e.g. Stan
- Today's presentation: BYM2 disconnected graphs (Streni-Ferrantino, 2018)

BYM2 model: Riebler et al, 2016

An intuitive Bayesian spatial model for disease mapping that accounts for scaling

https://arxiv.org/abs/1601.01180

Riebler innovation: put spatial component and random effects component on same scale

- BYM2 model where combination of spatial and non-spatial components $\phi + \theta$ is parameterized

$$\left((\sqrt{\rho/s}\ )\,\phi^* + (\sqrt{1-\rho})\,\theta^* \right)\sigma$$

  where:
  - $\sigma \geq 0$ is the overall standard deviation.
  - $\rho \in [0, 1]$ - proportion of spatial variance.
  - ϕ^* is the ICAR component (fully connected areal map)
  - $\theta^* \sim N(0, 1)$ is the vector of ordinary random effects
  - s is a scaling factor s.t. $\text{Var}(\phi_i) \approx 1$; s is data.

BYM2 for disconnected graphs: Freni-Sterrantino et al, 2017

A note on intrinsic conditional autoregressive models for disconnected graphs
https://arxiv.org/pdf/1705.04854.pdf

- scale each connected component of size larger than one as in BYM2 model
- impose sum-to-zero constraints on each connected component
- spatial variance for islands is standard Normal(0, 1)

Conceptually simple, tedious to implement

BYM2 connected, disconnected graphs in Stan

see Jupyter notebook