Unmeasured Spatial Risk in Disease Modeling

A Topological Data Analysis Approach

Hashan Fernando, Adam Spannaus, Anuj J. Kapadia, Jodie Trafton

Introduction

- Goal of this project is to predict opioid overdose risk using socioeconomic factors.
- For a county level disease model, we include census-tract level spatial insights.
- We use Topological Data Analysis (TDA) to identify and to quantify census-tract level spatial characteristics.
- Spatial characteristic help to estimate the disease risk more accurately.
- We use the topology of the Social Vulnerability Index (SVI) data to identify overdose risk.

Besag-York-Mollie 2 (BYM2) Model

- Deaths can be modeled as

$$\mathbf{y_i}|\theta_\mathbf{i} \sim \text{Poisson}(\mathbf{E_i}\theta_\mathbf{i}).$$

Where $\mathbf{y_i}$ is observed mortality count, $\mathbf{E_i}$ is expected number of deaths, and $\theta_\mathbf{i}$ is relative risk of an area i.

- Logarithm of relative risk($\eta_\mathbf{i}$) is defined as

$$\eta_\mathbf{i} = \mathbf{m} + \beta\mathbf{X_i^T} + \mathbf{b_i}.$$

Where m is intercept, β matrix is coefficients, $\mathbf{X^T}$ matrix is measurable covariates, $\mathbf{b_i}$ is spatial random effect.

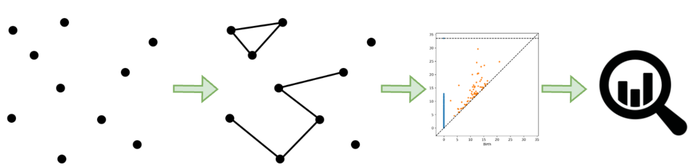
- Spatial random effect ($\mathbf{b_i}$)

$$\mathbf{b_i} = \frac{1}{\sqrt{\tau_\mathbf{b}}}(\sqrt{1-\phi}\mathbf{v_i} + \sqrt{\phi}\mathbf{u_{*i}}).$$

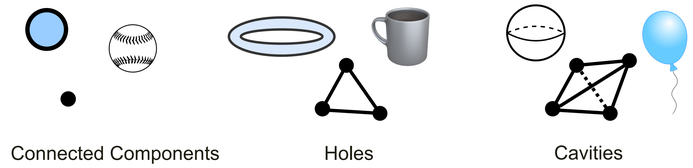
Where $\mathbf{u}$ is structured effect, $\mathbf{v}$ is unstructured effect, ϕ is mixing parameter, and $\tau_\mathbf{b}$ is precision parameter.

- The structured effect capture the spatial dependence due to neighboring regions, and the unstructured effect capture the random variation that cannot be explained by the rest of the model.
- A county level BYM-2 model is does not consider census level (within county) spatial characteristics. Therefore, we use TDA to quantify census level spatial characteristics of the census level SVI data.

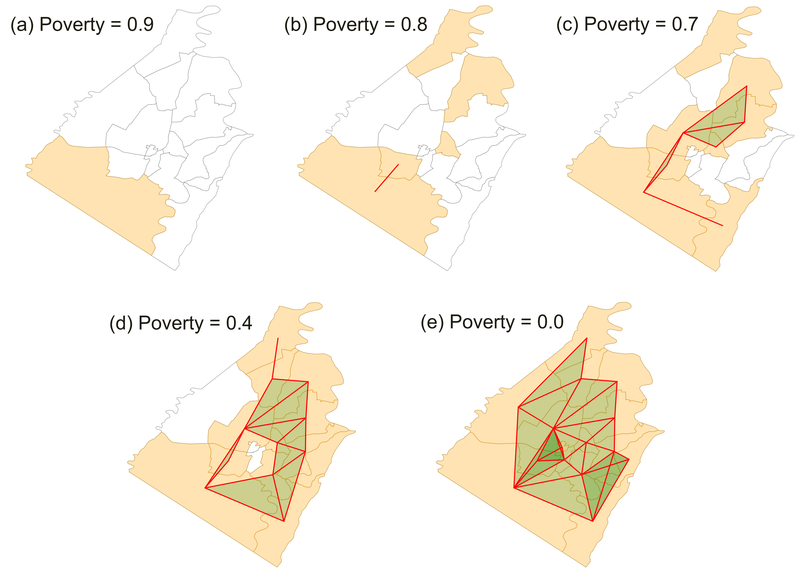
Topological Data Analysis



Topological Features

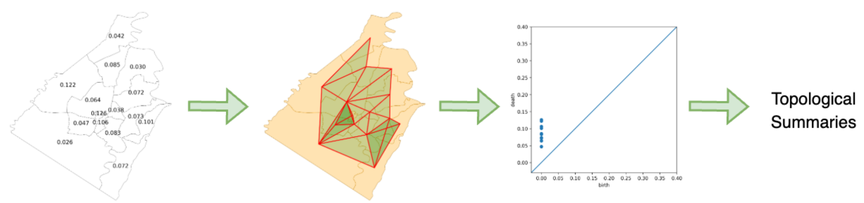


Geographical Adjacency Method



- Geographical adjacency complex is a type of a network constructed based on adjacency relationship between vertices(i.e., census tracts) and variable similarity.
- This method captures the topological structure of the data by connecting the vertices that are neighbors.
- In the method, we first sort the census-tracts based on the specific variable and then add one by one to the network by considering the order (fig (a) to (e)). If two counties that present in the network are adjacent, we create an edge.
- We employ geographic adjacency method to all the SVI variables to capture the spatial characteristics.

Data Processing Steps



- A persistence diagram is a collection of birth-death pairs that summarize the lifetimes of topological features in the data. We use these diagrams to capture and include topological feature information as part of our analysis.

BYM-T Model

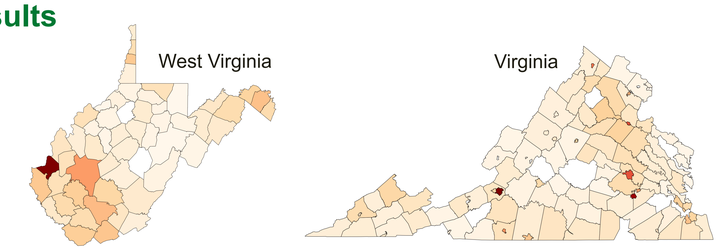
- BYM-T model is a new parameterization of BYM2 model.
- We calculate $\mathbf{s_i}$ by generating topological summaries for each variable and then aggregating them across all variables within region i.
- Key innovation here is, we extend the spatial component by adding combined topological summaries as $\mathbf{s_i}$. The spatial random effect can be denoted as

$$\mathbf{b_i} = \frac{1}{\sqrt{\tau_\mathbf{b}}}(\sqrt{\alpha_1}\mathbf{v_i} + \sqrt{\alpha_2}\mathbf{u_i} + \sqrt{\alpha_3}\mathbf{s_i}),$$

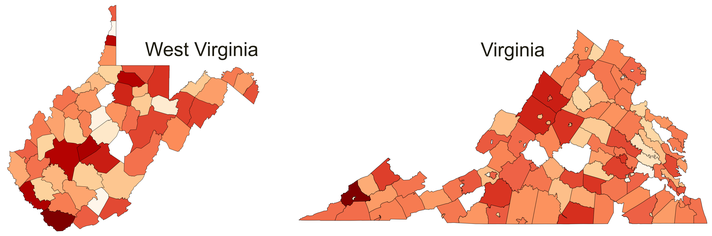
where

$$\alpha_1 + \alpha_2 + \alpha_3 = 1.$$

Results



The BYM2 Relative Risk Maps



Relative Risk Differences Between the BYM2 model and the BYM-T Model.

- In both West Virginia and Virginia, the BYM-T model outperformed base BYM2 model.
- Relative risk differences indicate that the BYM2 model tends to underestimate opioid overdose risk, while the BYM-T model captures additional overdose risk by incorporating census-level spatial characteristics.

Conclusions and Future Directions

- We successfully developed a framework that integrates topological features into spatial disease modeling.
- The results highlight that topological summaries improves risk estimates.
- In future work, we plan to incorporate a temporal component to extend our framework to predict the future opioid overdose risk.

References:

– Feng, Michelle, and Mason A. Porter. "Persistent homology of geospatial data: A case study with voting." *SIAM Review* 63.1 (2021): 67-99.

For more examples and code, check out our code repository:

Acknowledgements

This work is sponsored by the US Department of Veterans Affairs. This manuscript has been authored by UT-Battelle, LLC, under contract DE-AC05-00OR22725 with the US Department of Energy (DOE). The US government retains and the publisher, by accepting the article for publication, acknowledges that the US government retains a nonexclusive, paid-up, irrevocable, worldwide license to publish or reproduce the published form of this manuscript, or allow others to do so, for US government purposes. DOE will provide public access to these results of federally sponsored research in accordance with the DOE Public Access Plan (http://energy.gov/downloads/doe-public-access-plan).