Shrinkage in Space

Spillovers and Networks in a Hierarchical Model

Nikolas Kuschnig

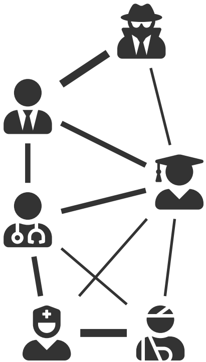
Causal Panel Data Conference, Stanford GSB

October 20th, 2023

Vienna University of Economics and Business
nikolas.kuschnig@wu.ac.at

Motivation

Economic activities rarely occur in isolation — agents are **embedded in networks** and **experience spillovers**.[a]



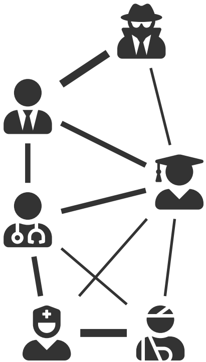
[a] See, e.g., Akcigit et al., 2021; Alfaro-Ureña et al., 2022; Ambrus and Elliott, 2021; Canen et al., 2023; Chetty et al., 2022; Dhyne et al., 2021; Giovanni et al., 2022; Vom Lehn and Winberry, 2022; Weidmann and Deming, 2021.

Motivation

Economic activities rarely occur in isolation — agents are **embedded in networks** and **experience spillovers.**[a]

The issue

We **rarely observe** the networks behind spillovers, and models suffer from the **curse of dimensionality**.



[a] See, e.g., Akcigit et al., 2021; Alfaro-Ureña et al., 2022; Ambrus and Elliott, 2021; Canen et al., 2023; Chetty et al., 2022; Dhyne et al., 2021; Giovanni et al., 2022; Vom Lehn and Winberry, 2022; Weidmann and Deming, 2021.

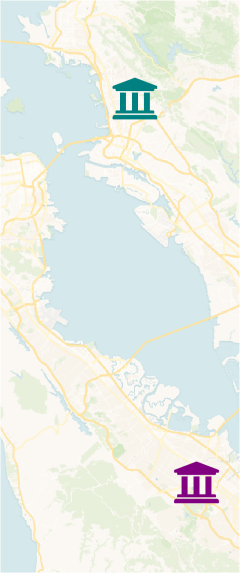
Overview

With networks unknown, models rely on **assumptions** and **approximate information**.

*How far is **Berkeley** from **Stanford**?*

Who are your five best friends?

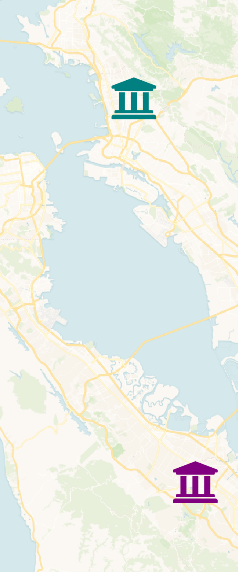
Who do you ask for advice?



Overview

With networks unknown, models rely on **assumptions** and **approximate information**.

> *How far is Berkeley from Stanford?*
>
> *Who are your five best friends?*
>
> *Who do you ask for advice?*



Today, I will show

- that these restrictions **distort inference**, and
- how to address this with a **Bayesian approach**.

Contributions and literature

Today, I'll focus on the *main contribution* to a growing literature[a] — a **Bayesian hierarchical approach** to model **spillovers** and **latent networks** behind them.

[a] Including Boucher and Houndetoungan, 2023; Debarsy and LeSage, 2022; de Paula et al., 2023; Goldsmith-Pinkham and Imbens, 2013; Griffith, 2022; Herstad, 2023; Hsieh and Lee, 2016; Lewbel et al., 2023; Zhang and Yu, 2018.

Contributions and literature

Today, I'll focus on the *main contribution* to a growing literature[a] — a **Bayesian hierarchical approach** to model **spillovers** and **latent networks** behind them.

Compared to the literature, my approach

- **flexibly leverages information** of all kinds,
- **naturally conveys uncertainty** via full posteriors,
- **is generally applicable.**

[a] Including Boucher and Houndetoungan, 2023; Debarsy and LeSage, 2022; de Paula et al., 2023; Goldsmith-Pinkham and Imbens, 2013; Griffith, 2022; Herstad, 2023; Hsieh and Lee, 2016; Lewbel et al., 2023; Zhang and Yu, 2018.

Information may include *geography*, *characteristics*, *group structures*, *proxies*, *repeated observations*, *sparsity*, *etc.* and is imposed via *structure* and *priors*.

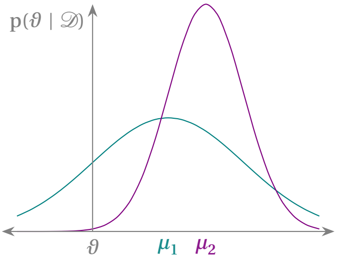
Contributions and literature

Today, I'll focus on the *main contribution* to a growing literature[a] — a **Bayesian hierarchical approach** to model **spillovers** and **latent networks** behind them.

Compared to the literature, my approach

- **flexibly leverages information** of all kinds,
- **naturally conveys uncertainty** via full posteriors,
- is generally applicable.

[a] Including Boucher and Houndetoungan, 2023; Debarsy and LeSage, 2022; de Paula et al., 2023; Goldsmith-Pinkham and Imbens, 2013; Griffith, 2022; Herstad, 2023; Hsieh and Lee, 2016; Lewbel et al., 2023; Zhang and Yu, 2018.



Information may include *geography, characteristics, group structures, proxies, repeated observations, sparsity, etc.* and is imposed via *structure* and *priors*.

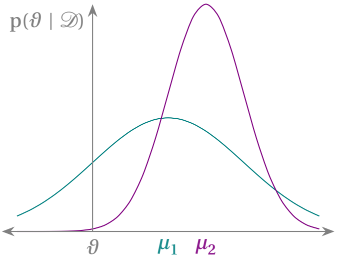
Contributions and literature

Today, I'll focus on the *main contribution* to a growing literature[a] — a **Bayesian hierarchical approach** to model **spillovers** and **latent networks** behind them.

Compared to the literature, my approach

- **flexibly leverages information** of all kinds,
- **naturally conveys uncertainty** via full posteriors,
- is **generally applicable**.

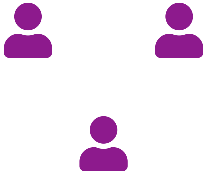
[a] Including Boucher and Houndetoungan, 2023; Debarsy and LeSage, 2022; de Paula et al., 2023; Goldsmith-Pinkham and Imbens, 2013; Griffith, 2022; Herstad, 2023; Hsieh and Lee, 2016; Lewbel et al., 2023; Zhang and Yu, 2018.



Information may include *geography*, *characteristics*, *group structures*, *proxies*, *repeated observations*, *sparsity*, *etc.* and is imposed via *structure* and *priors*.

Setting

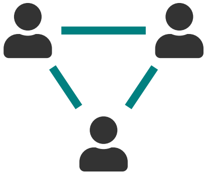
Consider a **set of agents** $\mathscr{A}$, for who we observe random *responses* $Y \in \mathbb{R}$ and *characteristics* $X \in \mathbb{R}^p$.



Setting

Consider a **set of agents** $\mathscr{A}$, for who we observe random *responses* $Y \in \mathbb{R}$ and *characteristics* $X \in \mathbb{R}^p$.

These agents have a **set of links** $\mathscr{E}$ between them — they are *connected in the network* $\mathscr{G} = \{\mathscr{A}, \mathscr{E}\}$.



Setting

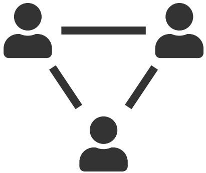
Consider a **set of agents** $\mathscr{A}$, for who we observe random *responses* $Y \in \mathbb{R}$ and *characteristics* $X \in \mathbb{R}^p$.

These agents have a **set of links** $\mathscr{E}$ between them — they are *connected in the network* $\mathscr{G} = \{\mathscr{A}, \mathscr{E}\}$.

We want to **learn about the relationship**

$$Y = f(X, \mathscr{G}) + \varepsilon,$$

and will need to *impose some structure* on f and $\mathscr{G}$.

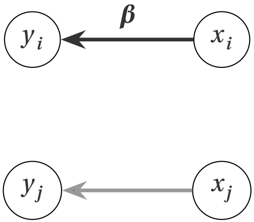


A model for f

Any economist's favorite model for f is

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{e}. \tag{1}$$

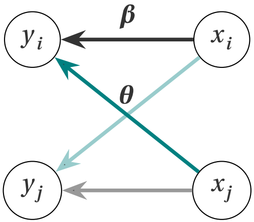
However, an agent's response may **depend on $\mathscr{G}$**.



A model for f

Any economist's favorite model for f is

$$\mathbf{y} = \mathbf{WX}\boldsymbol{\theta} + \mathbf{X}\boldsymbol{\beta} + \mathbf{e}. \tag{1}$$



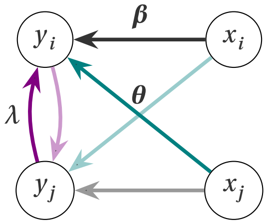
However, an agent's response may **depend on** $\mathscr{G}$.

- in terms of their **peers' characteristics**,

A model for f

Any economist's favorite model for f is

(1) $$\mathbf{y} = \lambda \mathbf{W}\mathbf{y} + \mathbf{W}\mathbf{X}\boldsymbol{\theta} + \mathbf{X}\boldsymbol{\beta} + \mathbf{e}.$$



However, an agent's response may **depend on** $\mathscr{G}$, e.g.,

- in terms of their **peers' characteristics**,
- and the **responses of their peers**.

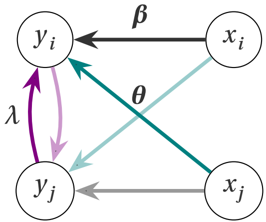
A model for f

Any economist's favorite model for f is

(1) $$\mathbf{y} = \lambda \mathbf{W}\mathbf{y} + \mathbf{W}\mathbf{X}\boldsymbol{\theta} + \mathbf{X}\boldsymbol{\beta} + \mathbf{e}.$$

However, an agent's response may **depend on** $\mathscr{G}$, e.g.,

- in terms of their **peers' characteristics**,
- and the **responses of their peers**.



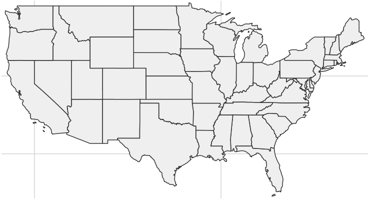
Linear network model

The network is represented by $\mathbf{W}$. Special cases are the *linear-in-means* and *spatial Durbin* models, which constrain $\mathbf{W}$ and treat it as given.

Does the network matter?

Consider *network effects*[a] based on

1. **contiguity of US states**, proxied with
2. **averages of contiguous states**, and
3. **distance-decay** between centers.

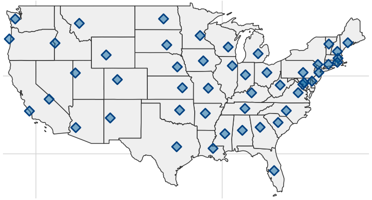


[a] The true values are $\lambda = 0.3, \theta = 0.5$.

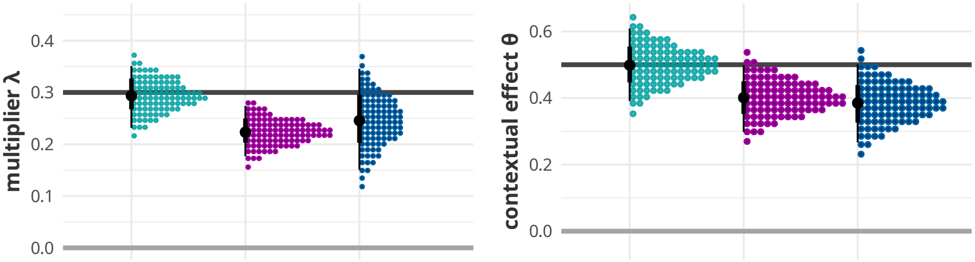
Does the network matter?

Consider *network effects*[a] based on

1. **contiguity of US states**, proxied with
2. **averages of contiguous states**, and
3. **distance-decay** between centers.



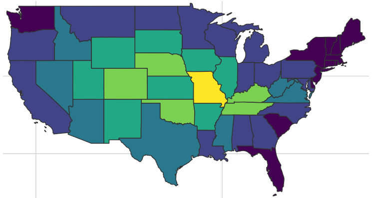
[a] The true values are $\lambda = 0.3, \theta = 0.5$.



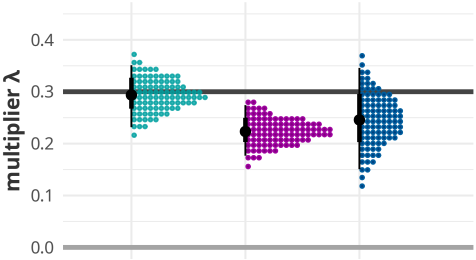
Does the network matter?

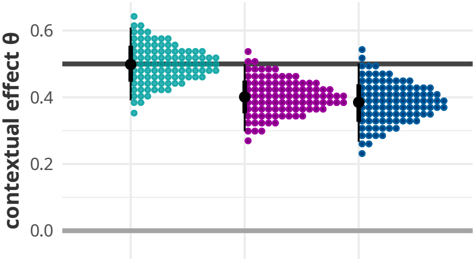
Consider *network effects*[a] based on

1. **contiguity of US states**, proxied with
2. **averages of contiguous states**, and
3. **distance-decay** between centers.



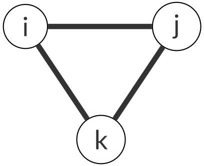
[a] The true values are $\lambda = 0.3, \theta = 0.5$.





The network

We will represent the network with the **graph** $\mathscr{G} = \{\mathscr{A}, \mathscr{E}\}$, which we allow to be

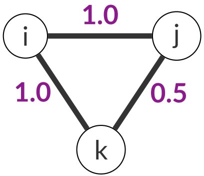


The network

We will represent the network with the **graph** $\mathscr{G} = \{\mathscr{A}, \mathscr{E}\}$, which we allow to be

- **weighted** — links are induced and measured by

$$g : \mathscr{A} \times \mathscr{A} \mapsto \mathbb{R}^{+},$$



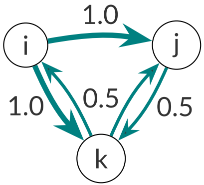
The network

We will represent the network with the **graph** $\mathscr{G} = \{\mathscr{A}, \mathscr{E}\}$, which we allow to be

- **weighted** — links are induced and measured by

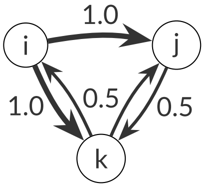
$$g : \mathscr{A} \times \mathscr{A} \mapsto \mathbb{R}^+,$$

- **directed** — links need not be reciprocal.



The network

We will represent the network with the **graph** $\mathscr{G} = \{\mathscr{A}, \mathscr{E}\}$, which we allow to be



- **weighted** — links are induced and measured by

$$g : \mathscr{A} \times \mathscr{A} \mapsto \mathbb{R}^+,$$

- **directed** — links need not be reciprocal.

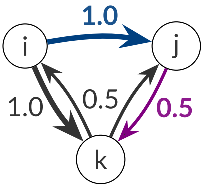
Adjacency matrix

The graph corresponds to the matrix $\mathbf{G}$ with entries given by $g_{ij} = g(i, j)$.

$$\mathbf{G} = \begin{bmatrix} 0 & g_{12} & \cdots & g_{1n} \\ g_{21} & 0 & \cdots & g_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ g_{n1} & g_{n2} & \cdots & 0 \end{bmatrix}.$$

The normalized adjacency matrix

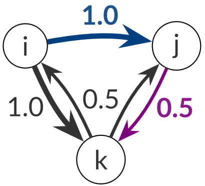
In practice, a **normalized adjacency matrix $\mathbf{W}$** is used, such that λ and $\boldsymbol{\theta}$ are identified.



$$\mathbf{G} = \begin{bmatrix} 0 & 1.0 & 1.0 \\ 0 & 0 & 0.5 \\ 0.5 & 0.5 & 0 \end{bmatrix},$$

The normalized adjacency matrix

In practice, a **normalized adjacency matrix $\mathbf{W}$** is used, such that λ and $\boldsymbol{\theta}$ are identified.



Row normalization

The standard is to transform $\tilde{\mathbf{W}}$ to be **row-stochastic**, such that $\sum_j w_{ij} = 1 \; \forall i$.

$$\mathbf{G} = \begin{bmatrix} 0 & 1.0 & 1.0 \\ 0 & 0 & 0.5 \\ 0.5 & 0.5 & 0 \end{bmatrix},$$

$$\tilde{\mathbf{W}} = \begin{bmatrix} 0 & 0.5 & 0.5 \\ 0 & 0 & 1.0 \\ 0.5 & 0.5 & 0 \end{bmatrix}.$$

The normalized adjacency matrix

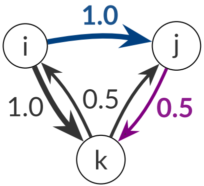
In practice, a **normalized adjacency matrix W** is used, such that λ and $\boldsymbol{\theta}$ are identified.

Row normalization

The standard is to transform $\tilde{\mathbf{W}}$ to be **row-stochastic**, such that $\sum_j w_{ij} = 1 \; \forall i$.

Scalar normalization

We will use **scalar normalization**, such that $w_{ij} = g_{ij} \times \varsigma \; \forall i, j$, in order to **preserve the network structure.** See more



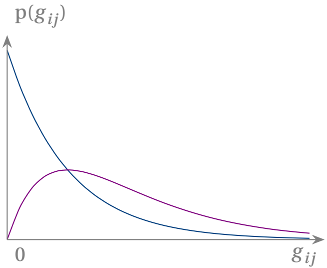
$$\mathbf{G} = \begin{bmatrix} 0 & \boxed{1.0} & 1.0 \\ 0 & 0 & \boxed{0.5} \\ 0.5 & 0.5 & 0 \end{bmatrix} = \boxed{\mathbf{W}},$$

$$\tilde{\mathbf{W}} = \begin{bmatrix} 0 & \boxed{0.5} & 0.5 \\ 0 & 0 & \boxed{1.0} \\ 0.5 & 0.5 & 0 \end{bmatrix}.$$

Network model — full parameterization

We want to model links, and could do so directly

$$g_{ij} \sim f(\cdot) \quad \forall i \neq j.$$



Network model — full parameterization

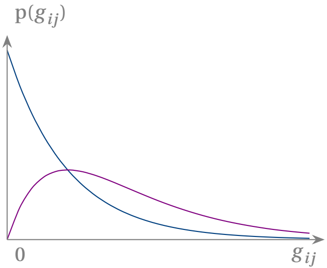
We want to model links, and could do so directly

$$g_{ij} \sim f(\cdot) \quad \forall i \neq j.$$

At $\mathcal{O}\left(N^2\right)$ **unknown links**, we'd need either

- **repeated observations** of the network, or
- **heavy shrinkage** to make this work.

Essentially, this is the approach of de Paula et al. (2023), who regularize using an *elastic net*.



Network model — full parameterization

We want to model links, and could do so directly

$$g_{ij} \sim f(\cdot) \quad \forall i \neq j.$$

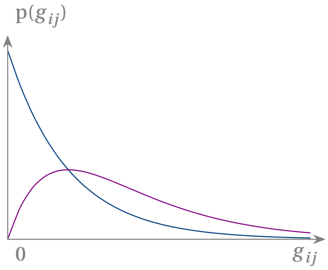
At $\mathcal{O}(N^2)$ **unknown links**, we'd need either

- **repeated observations** of the network, or
- **heavy shrinkage** to make this work.

We want to **constrain the dimensionality** by **imposing some structure**[a] on $\mathcal{G}$, allowing for *more nuance where it is needed.*

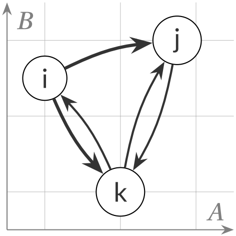
[a] Lewbel et al. (2023), e.g., constrain links to *sub-networks*. ▸ Illustration

Essentially, this is the approach of de Paula et al. (2023), who regularize using an *elastic net*.



Imposing structure — A (Metric) Space Odyssey

Alternatively, assume that we can **locate our agents** in some (generalized) **metric space** $(\mathscr{P}, \mathrm{d})$.[a]

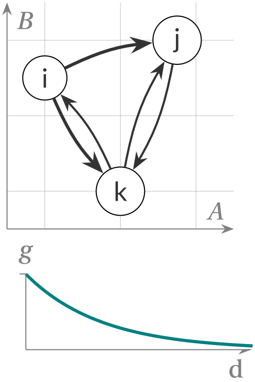


[a] This is rather natural in a spatial setting, and has been used successfully used for modelling social networks (going back to Hoff et al., 2002).

Alternatively, assume that we can **locate our agents** in some (generalized) **metric space** $(\mathscr{P}, \mathrm{d})$.[a]

Then, we can think of **links as decaying in the distance** between *latent positions* $P \in \mathbb{R}^D$ of agents, e.g.

$$g_{ij} = \exp\left\{-\mathrm{d}_{ij}\right\} \quad \forall i \neq j.$$



[a] This is rather natural in a spatial setting, and has been used successfully used for modelling social networks (going back to Hoff et al., 2002).

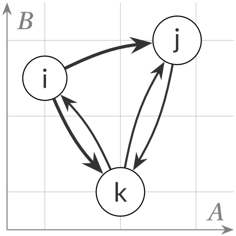
Imposing structure — A (Metric) Space Odyssey

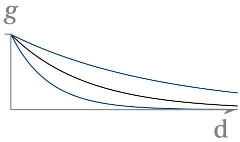
Alternatively, assume that we can **locate our agents** in some (generalized) **metric space** $(\mathscr{P}, \mathrm{d})$.[a]

Then, we can think of **links as decaying in the distance** between *latent positions* $P \in \mathbb{R}^D$ of agents, e.g.

$$g_{ij} = \exp\left\{-\delta \times \mathrm{d}_{ij}\right\}.$$

- We may also consider, e.g., the **speed of decay**,





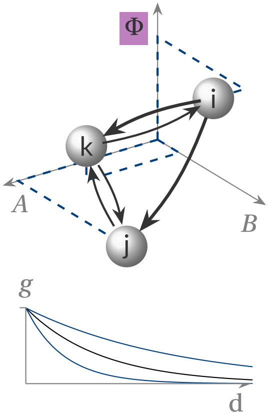
[a] This is rather natural in a spatial setting, and has been used successfully used for modelling social networks (going back to Hoff et al., 2002).

Imposing structure — A (Metric) Space Odyssey

Alternatively, assume that we can **locate our agents** in some (generalized) **metric space** $(\mathscr{P}, \mathrm{d})$.[a]

Then, we can think of **links as decaying in the distance** between *latent positions* $P \in \mathbb{R}^D$ of agents, e.g.

$$g_{ij} = \exp\left\{-\boxed{\delta} \times \boxed{\phi_i^{-1}} \times \boxed{\mathrm{d}_{ij}}\right\}.$$



- We may also consider, e.g., the **speed of decay**,
- or *asymmetries* via **popularity** or gravity.

[a]This is rather natural in a spatial setting, and has been used successfully used for modelling social networks (going back to Hoff et al., 2002).

$\mathbf{y} = \lambda \mathbf{W}\mathbf{y} + \dot{\mathbf{W}}\mathbf{X}\boldsymbol{\theta} + \mathbf{X}\boldsymbol{\beta} + \mathbf{e}$, where $\mathbf{W}, \dot{\mathbf{W}} = f(\cdot)$

$\mathbf{y} = \lambda \mathbf{W}\mathbf{y} + \dot{\mathbf{W}}\mathbf{X}\boldsymbol{\theta} + \mathbf{X}\boldsymbol{\beta} + \mathbf{e}, \text{ where } \mathbf{W}, \dot{\mathbf{W}} = f(\cdot)$

Nested specifications

Latent positions may be informed by *geographical coordinates*, by *homophilic characteristics*, or entirely unknown.

$\mathbf{y} = \lambda \mathbf{W}\mathbf{y} + \dot{\mathbf{W}}\mathbf{X}\boldsymbol{\theta} + \mathbf{X}\boldsymbol{\beta} + \mathbf{e}, \text{ where } \mathbf{W}, \dot{\mathbf{W}} = f(\cdot)$

Nested specifications

Latent positions may be informed by *geographical coordinates*, by *homophilic characteristics*, or entirely unknown.

Flexibility

Depending on the *setting* and available information, we **adjust the structure** and fix, shrink, or free up **parameters**.

$\mathbf{y} = \lambda \mathbf{W}\mathbf{y} + \dot{\mathbf{W}}\mathbf{X}\boldsymbol{\theta} + \mathbf{X}\boldsymbol{\beta} + \mathbf{e}, \text{ where } \mathbf{W}, \dot{\mathbf{W}} = f(\cdot)$

Nested specifications

Latent positions may be informed by *geographical coordinates*, by *homophilic characteristics*, or entirely unknown.

Flexibility

Depending on the *setting* and available information, we **adjust the structure** and fix, shrink, or free up **parameters**.

Estimation or: How I Learned to Stop Worrying and Love MCMC

Adaptive MCMC facilitates full posterior inference, nuanced **weakly informative priors** improve convergence, and a **Gaussian process approximation** for costly Jacobians improves speed. ▸ See more

The approach in practice

We'll simulate repeatedly from

$$\mathbf{y}_t = \alpha + \lambda \mathbf{W}\mathbf{y}_t + \mathbf{x}_t\beta + \dot{\mathbf{W}}\mathbf{x}_t\theta + \mathbf{e}_t, \text{ where } \mathbf{x}_t, \mathbf{e}_t \sim \mathrm{N}(0, 1).$$

The networks behind $\mathbf{W}$ stem from

1. distance between **US population centers**
   - sparse/dense, asymmetric/symmetric
2. random **Erdős–Rényi graph** — $\mathscr{M}_{\text{open}}$.

The approach in practice

We'll simulate repeatedly from

$$\mathbf{y}_t = \alpha + \lambda \mathbf{W}\mathbf{y}_t + \mathbf{x}_t\beta + \dot{\mathbf{W}}\mathbf{x}_t\theta + \mathbf{e}_t, \text{ where } \mathbf{x}_t, \mathbf{e}_t \sim \mathrm{N}(0, 1).$$

The networks behind $\mathbf{W}$ stem from

1. distance between **US population centers**
   - sparse/dense, asymmetric/symmetric
2. random **Erdős–Rényi graph** — $\mathscr{M}_{\text{open}}$.

Setup

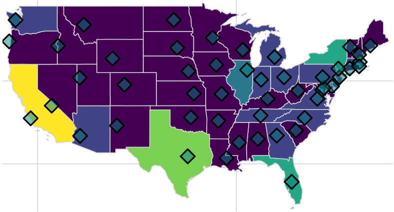
I'll model the *locations*, speed of *distance-decay*, and *popularity*. The **priors will be too flat**, and the *sampler's initialised* first at the true values, then at draws from the prior.

Simulation results — US population

The first network is determined as

$$g_{ij} = \exp\left\{-\delta_i \times \mathrm{d}\left(\mathbf{p}_i, \mathbf{p}_j\right)\right\},$$

where **popularity** δ_i is driven by a state's population, and $\mathbf{p}_i$ its **population center**.

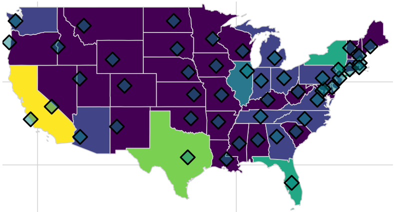


Simulation results — US population

The first network is determined as

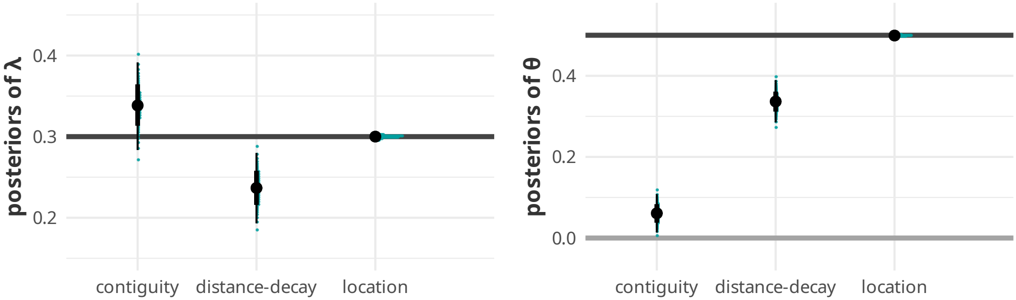
$$g_{ij} = \exp\left\{-\delta_i \times \mathrm{d}\left(\mathbf{p}_i, \mathbf{p}_j\right)\right\},$$

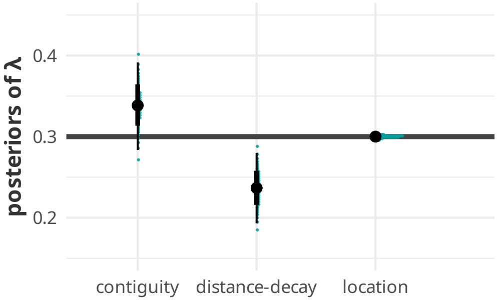
where **popularity** δ_i is driven by a state's population, and $\mathbf{p}_i$ its **population center**.

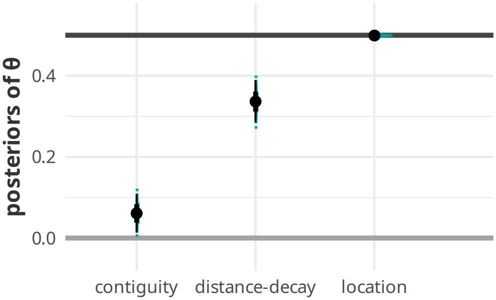


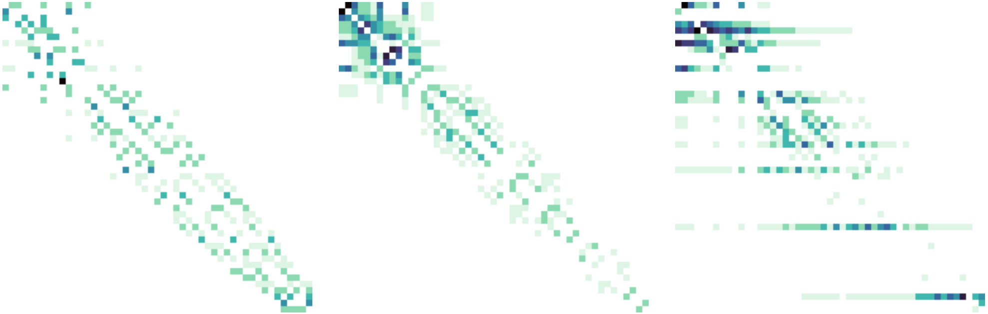
I'll start by estimating λ, θ in a long panel ($T = 50$) using

- **contiguity** between states,
- **distance-decay** between centers, and
- the *true model* of **location & popularity**.


posteriors of λ
0.4
0.3
0.2
contiguity
distance-decay
location
posteriors of θ
0.4
0.2
0.0
contiguity
distance-decay
location







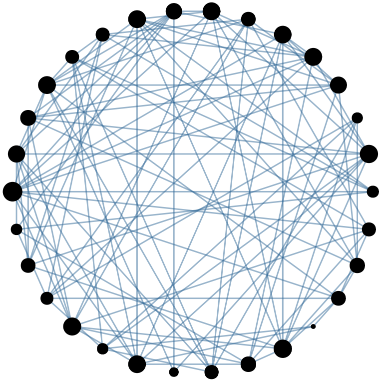
W based on **contiguity**, **distance-decay**, and the **true network**.

Simulation results — the Erdős–Rényi graph

The graph ($N \in \{30, 50\}$) is determined as

$$g_{ij} = \begin{cases} 1 \text{ with probability } 0.25, \\ 0 \text{ otherwise.} \end{cases}$$

Our model uses the same **distance-decay** specification as before.



A realization of $\mathscr{G}(30, 0.25)$.

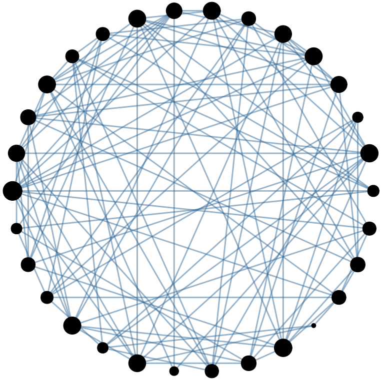
Simulation results — the Erdős–Rényi graph

The graph ($N \in \{30, 50\}$) is determined as

$$g_{ij} = \begin{cases} 1 \text{ with probability } 0.25, \\ 0 \text{ otherwise.} \end{cases}$$

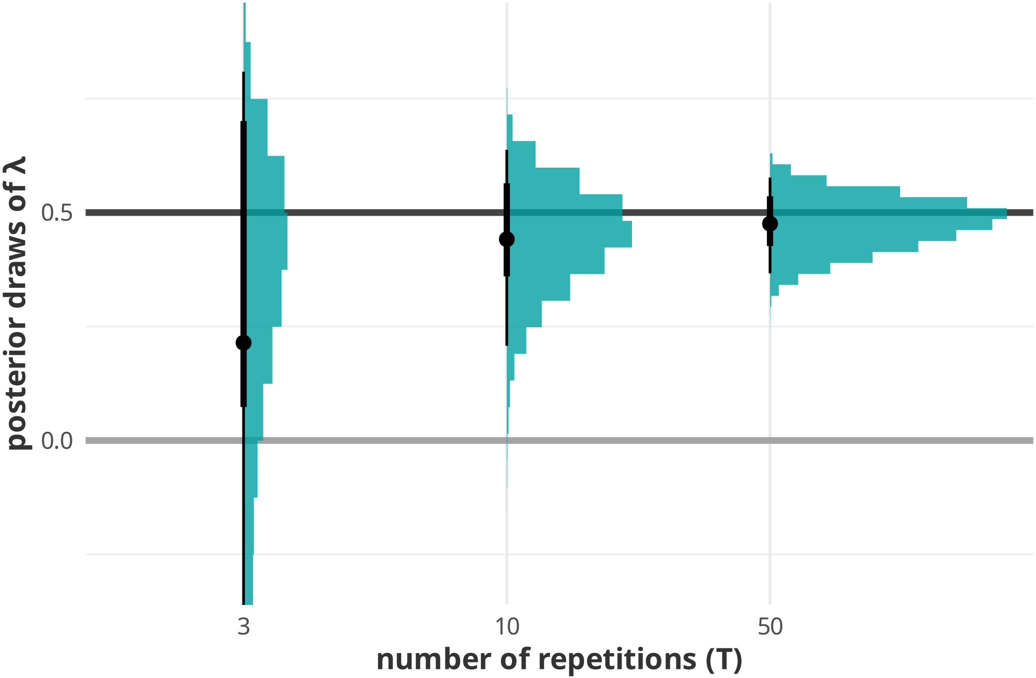
Our model uses the same **distance-decay** specification as before.

To **highlight convergence**, we'll have a look at **posteriors of λ** for multiple simulations after a **short burn-in** of 1,000 draws.

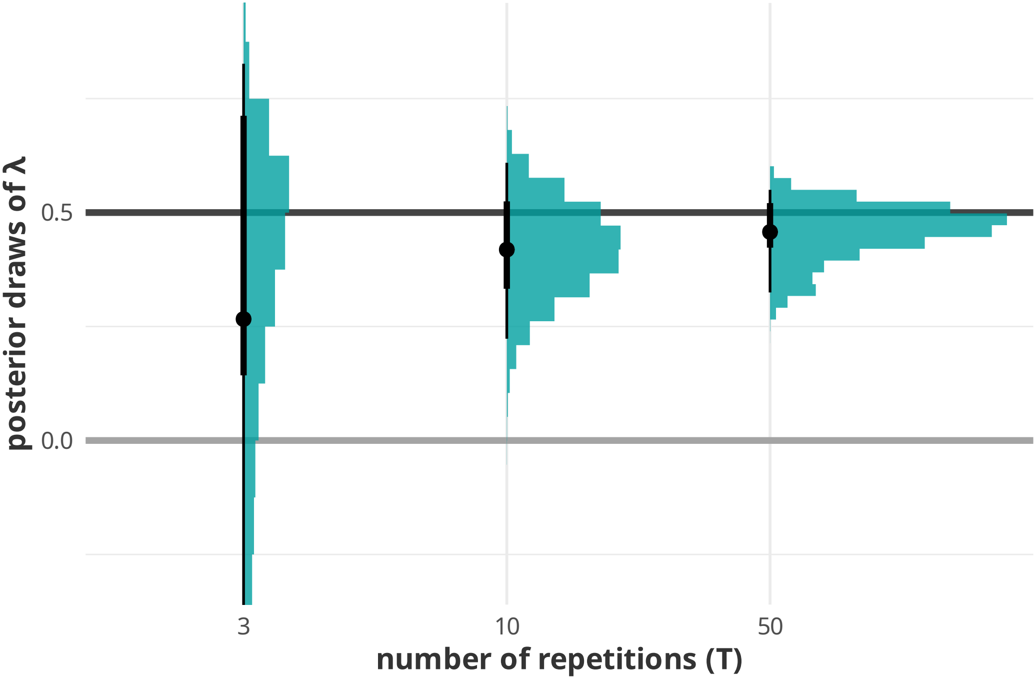


A realization of $\mathcal{G}(30, 0.25)$.

Connectivity strength λ (N = 30)



Connectivity strength λ (N = 50)



Conclusion

- I developed a **framework for jointly modelling** f and $\mathscr{G}$,
- that flexibly leverages **data**, **structure**, and **shrinkage**.
- It's widely applicable to **network and spatial settings**
  - with *no, limited, or uncertain information* on *diverse links*,
- allowing us to gain **deeper insights into spillovers**,
  - at moderate to pronounced *computational costs*.

Conclusion

- I developed a **framework for jointly modelling** f and $\mathscr{G}$,
- that flexibly leverages **data**, **structure**, and **shrinkage**.
- It's widely applicable to **network and spatial settings**
    - with *no, limited, or uncertain information* on *diverse links*,
- allowing us to gain **deeper insights into spillovers**,
    - at moderate to pronounced *computational costs*.

Conclusion

- I developed a **framework for jointly modelling** f and $\mathscr{G}$,
- that flexibly leverages **data**, **structure**, and **shrinkage**.
- It's widely applicable to **network and spatial settings**
  - with *no*, *limited*, *or uncertain information* on *diverse links*,
- allowing us to gain **deeper insights into spillovers**,
  - at moderate to pronounced *computational costs*.

Conclusion

- I developed a **framework for jointly modelling** f and $\mathscr{G}$,
- that flexibly leverages **data**, **structure**, and **shrinkage**.
- It's widely applicable to **network and spatial settings**
  - with *no*, *limited, or uncertain information* on *diverse links*,
- allowing us to gain **deeper insights into spillovers**,
  - at moderate to pronounced *computational costs*.

Conclusion

- I developed a **framework for jointly modelling** f and $\mathscr{G}$,
- that flexibly leverages **data**, **structure**, and **shrinkage**.
- It's widely applicable to **network and spatial settings**
  - with *no*, *limited*, *or uncertain information* on *diverse links*,
- allowing us to gain **deeper insights into spillovers**,
  - at moderate to pronounced *computational costs*.

For **more details and info** on identification, priors, sampling, and applications have a look at *the appendix*, or — coming soon™ to a repository near you — a draft.

References i

Akcigit, Ufuk, Douglas Hanley, and Nicolas Serrano-Velarde (2021). **"Back to basics: basic research spillovers, innovation policy, and growth".** In: *Review of Economic Studies* 88.1, pp. 1–43. ISSN: 0034-6527. DOI: `10.1093/restud/rdaa061`.

Alfaro-Ureña, Alonso, Isabela Manelici, and Jose P. Vasquez (2022). **"The effects of joining multinational supply chains: New evidence from firm-to-firm linkages".** In: *Quarterly Journal of Economics* 137.3, pp. 1495–1552. ISSN: 0033-5533. DOI: `10.1093/qje/qjac006`.

Ambrus, Attila and Matt Elliott (2021). **"Investments in social ties, risk sharing, and inequality".** In: *Review of Economic Studies* 88.4, pp. 1624–1664. ISSN: 0034-6527. DOI: `10.1093/restud/rdaa073`.

Boucher, Vincent and Elysée Aristide Houndetoungan (2023). **"Estimating peer effects using partial network data".** In: *Working Paper*.

References ii

Canen, Nathan, Matthew O. Jackson, and Francesco Trebbi (2023). **"Social interactions and legislative activity".** In: *Journal of the European Economic Association* 21.3, pp. 1072–1118. ISSN: 1542-4766. DOI: `10.1093/jeea/jvac051`.

Chetty, Raj et al. (2022). **"Social capital I: measurement and associations with economic mobility".** In: *Nature* 608.7921, pp. 108–121. ISSN: 1476-4687. DOI: `10.1038/s41586-022-04996-4`.

de Paula, Áureo, Imran Rasul, and Pedro Souza (2023). ***Identifying network ties from panel data: Theory and an application to tax competition.*** DOI: `10.48550/arXiv.1910.07452`.

Debarsy, Nicolas and James P. LeSage (2022). **"Bayesian model averaging for spatial autoregressive models based on convex combinations of different types of connectivity matrices".** In: *Journal of Business & Economic Statistics*, pp. 1–12. DOI: `10.1080/07350015.2020.1840993`.

References iii

- Dhyne, Emmanuel et al. (2021). “**Trade and domestic production networks**”. In: *Review of Economic Studies* 88.2, pp. 643–668. ISSN: 0034-6527. DOI: 10.1093/restud/rdaa062.
- Giovanni, Julian di et al. (2022). “**International spillovers and local credit cycles**”. In: *Review of Economic Studies* 89.2, pp. 733–773. ISSN: 0034-6527. DOI: 10.1093/restud/rdab044.
- Goldsmith-Pinkham, Paul and Guido W. Imbens (2013). “**Social networks and the identification of peer effects**”. In: *Journal of Business & Economic Statistics* 31.3, pp. 253–264. DOI: 10.1080/07350015.2013.801251.
- Griffith, Alan (2022). “**Name your friends, but only five? The importance of censoring in peer effects estimates using social network data**”. In: *Journal of Labor Economics*. DOI: 10.1086/717935.
- Herstad, Eyo (2023). “**Estimating peer effects and network formation models with missing network links**”. In: *Working Paper*.

References iv

- Hoff, Peter D., Adrian E. Raftery, and Mark S. Handcock (2002). **"Latent space approaches to social network analysis".** In: *Journal of the American Statistical Association* 97.460, pp. 1090–1098. ISSN: 0162-1459. DOI: `10.1198/016214502388618906`.
- Hsieh, Chih-Sheng and Lung Fei Lee (2016). **"A social interactions model with endogenous friendship formation and selectivity".** In: *Journal of Applied Econometrics* 31.2, pp. 301–319. DOI: `10.1002/jae.2426`.
- Lewbel, Arthur, Xi Qu, and Xun Tang (2023). **"Social networks with unobserved links".** In: *Journal of Political Economy*. DOI: `10.1086/722090`.
- Vom Lehn, Christian and Thomas Winberry (2022). **"The investment network, sectoral comovement, and the changing U.S. business cycle".** In: *Quarterly Journal of Economics* 137.1, pp. 387–433. ISSN: 0033-5533. DOI: `10.1093/qje/qjab020`.
- Weidmann, Ben and David J. Deming (2021). **"Team players: How social skills improve team performance".** In: *Econometrica* 89.6, pp. 2637–2657. ISSN: 1468-0262. DOI: `10.3982/ECTA18461`.

Zhang, Xinyu and Jihai Yu (2018). "**Spatial weights matrix selection and model averaging for spatial autoregressive models**". In: *Journal of Econometrics* 203.1, pp. 1–18. DOI: 10.1016/j.jeconom.2017.05.021.

Model

We are interested in

$$p(\Theta, \mathscr{E} \mid \mathscr{D}) \propto p(\Theta, \mid \mathscr{E}, \mathscr{D}) \times p(\mathscr{E} \mid \Theta, \mathscr{D}),$$

or, to be more concrete, in

$$\mathbf{y} = (\mathbf{I} - \lambda \mathbf{W})^{-1} \mathbf{z},$$
$$\mathbf{z} = \mathbf{X}\boldsymbol{\beta} + \dot{\mathbf{W}}\mathbf{X}\boldsymbol{\theta} + \boldsymbol{\varepsilon},$$

where $\mathbf{W} = g(\cdot)\varsigma$, $\dot{\mathbf{W}} = g(\cdot)\dot{\varsigma}$, and g is based on a network model of choice. Options include the ones described, many others, or a combination thereof.

Identification

We can identify the parameters $\lambda, \boldsymbol{\theta}$ with mild constraints. Network parameters are generally only weakly identified. We can alleviate this by imposing constraints from the literature or prior information.

Normalization — multiplier effect

Consider a **network autoregression**

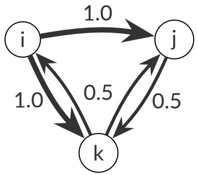
$$\mathbf{y} = \lambda \mathbf{W}\mathbf{y} + \mathbf{e}.$$

The following result guarantees stability.

Theorem

Let $\mathbf{I}$ denote the identity matrix, and α be a real scalar. Then $\mathbf{I} - \alpha\mathbf{A}$ is stationary for $\alpha \in (-\rho_{\mathbf{A}}, \rho_{\mathbf{A}})$, where $\rho_{\mathbf{A}}$ denotes the spectral radius of $\mathbf{A}$.

By normalizing with the *spectral radius* (i.e. $\varsigma = \rho_{\mathbf{G}}^{-1}$), we can let $\lambda \in (-1, 1)$. This relates λ to the *dominant eigenvector* of the network, which generally *does not coincide* with the *average partial effect* of $\mathbf{W}$.



Row-normalization distorts, e.g., the *eigenvector centrality* c.

	i	j	k
$c_{\mathbf{G}}$	2^{-1}	6^{-1}	3^{-1}
$c_{\tilde{\mathbf{W}}}$	3^{-1}	3^{-1}	3^{-1}.

In fact, λ is *at least* the average partial effect (by the spectral radius being the infimum norm).

Normalization — contextual effect

Consider a contextual model

$$\mathbf{y} = \lambda \dot{\mathbf{W}} \mathbf{X} \boldsymbol{\theta} + \mathbf{e}.$$

$$\frac{\partial \mathbf{z}}{\partial \mathbf{x}_k} = \mathbf{I}\beta_k + \dot{\mathbf{W}}\theta_k.$$

In this case, we have fewer requirements of the normalization. One sensible option is to fix $\boldsymbol{\theta}$ at the *average partial effect* of the network characteristics.[a] We can achieve that by setting $\dot{\mathbf{W}}$ such that it sums to N – we scale with $\dot{\varsigma} = \frac{N}{\sum_i \sum_j g_{ij}}$.
This applies similarly to the nested linear model when a network multiplier is present. ▸ Go back

[a] An alternative when some agents are not linked in the network, is the average partial effect for all agents that are linked within the network.

Estimation

For full posterior inference, we extend existing *MCMC methods* — the central term (suppressing contextual effects) is given by

$$|\mathbf{S}(\lambda,\cdot)| \exp\left\{-\frac{1}{2\sigma^2}\left(\mathbf{S}(\lambda,\cdot)\mathbf{y}-\mathbf{X}\boldsymbol{\beta}\right)'\left(\mathbf{S}(\lambda,\cdot)\mathbf{y}-\mathbf{X}\boldsymbol{\beta}\right)\right\},$$

where $\mathbf{S}(\lambda,\cdot) = (\mathbf{I}-\lambda\mathbf{W})$ is a spatial filter. The main concerns are essentially computational — we need

1. convergence of parameters,
2. to evaluate the $N \times N$ **Jacobian determinant**.

We use *Gibbs* and *Metropolis-Hastings* steps, as well as a *rejection sampler* to draw from the posterior.

Weakly informative priors

We use weakly informative priors to (1) help improve convergence of the MCMC samples, and (2) build a more credible, realistic model.

Two central parameters are λ and $\boldsymbol{\theta}$. For the former, I propose the *hierarchical prior*

$$\lambda \sim \text{Beta}\left(1+\tau, 1+\tau\right), \quad \tau \sim \text{Gamma},$$

which adds barely any computational overhead,[a] but facilitates much better shrinkage towards, e.g., *zero*, while providing wide support.

For $\boldsymbol{\theta}$, which resembles a standard coefficient, standard global-local shrinkage priors are applicable.

[a] Thanks to a *rejection sampler* based on a Gamma proposal density.

Parameters for the network structure also greatly benefit from weakly informative priors, and even more from actually informative ones. Options for $\boldsymbol{\theta}$ include the Horseshoe, Dirichlet–Laplace, and the Normal-Gamma shrinkage priors.

A Beta prior ...

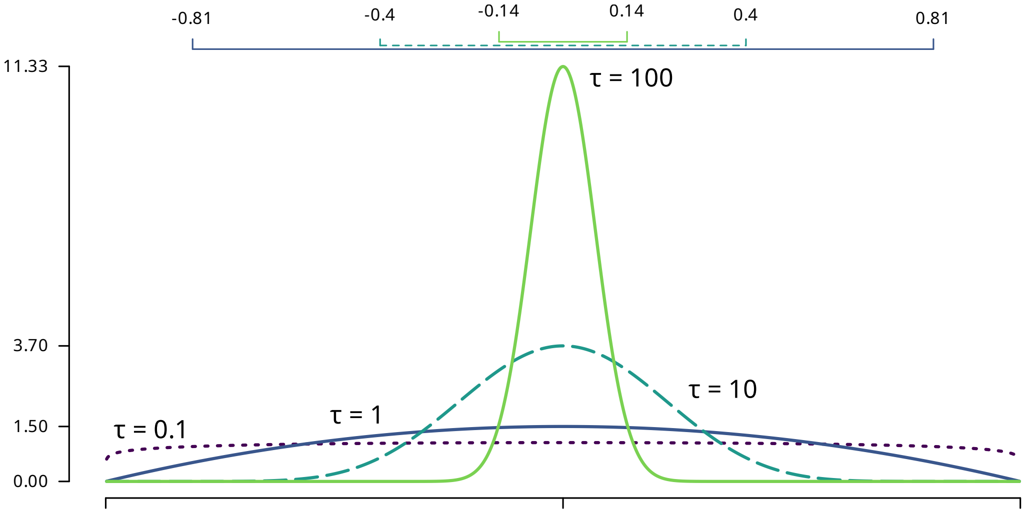


Figure 1: Scaled $\mathrm{Beta}(1 + \tau, 1 + \tau)$ densities with increasing weight, τ.

...with a Gamma mixing distribution

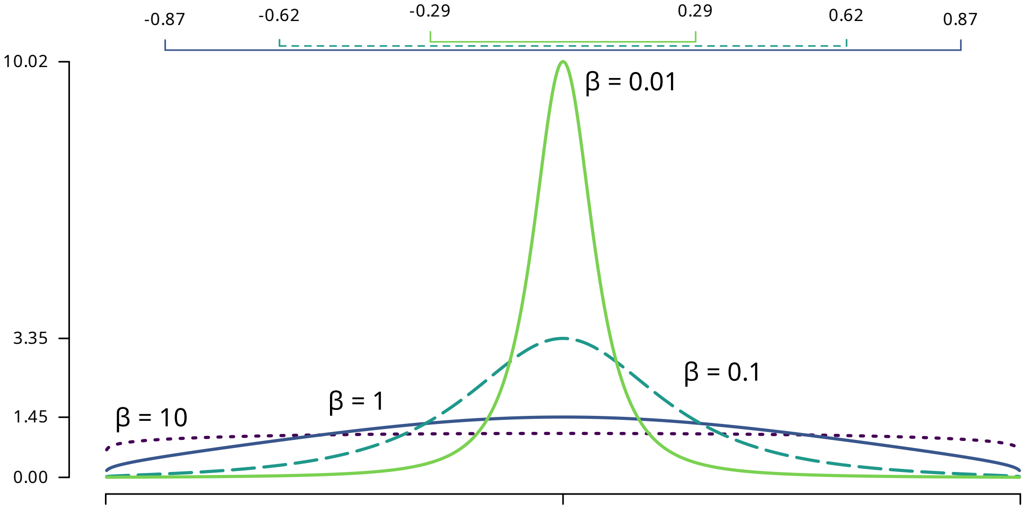


Figure 2: Scaled $\mathrm{Beta}(1+\tau, 1+\tau)$, $\tau \sim \mathrm{Exp}(\beta)$ with increasing weight, $\mathbb{E}[\tau] = 1/\beta$.

Go back

Jacobian determinant

In standard models, we'd compute the **spectral decomposition** of $\mathbf{W}$ **once** and compute the determinant using $\mathbf{W}$'s eigenvalues (η_i), as

$$\ln |\mathbf{I} - \lambda \mathbf{W}| = \sum_{i=1}^{N} \ln \left(1 - \lambda \eta_i\right).$$

However, our $\mathbf{W}$ is **mutable**, and computing eigenvalues for every draw of δ is prohibitive at $\mathcal{O}(N^3)$ complexity.

For models with limited parameters for the network structur, I propose a **Gaussian process approximation** instead –

$$\left|\mathbf{S}(\lambda, \cdot)\right| \approx \mathrm{GP}\left(\mu(\lambda, \cdot), \mathbf{\Sigma}(\lambda, \cdot)\right).$$

Gaussian process approximation

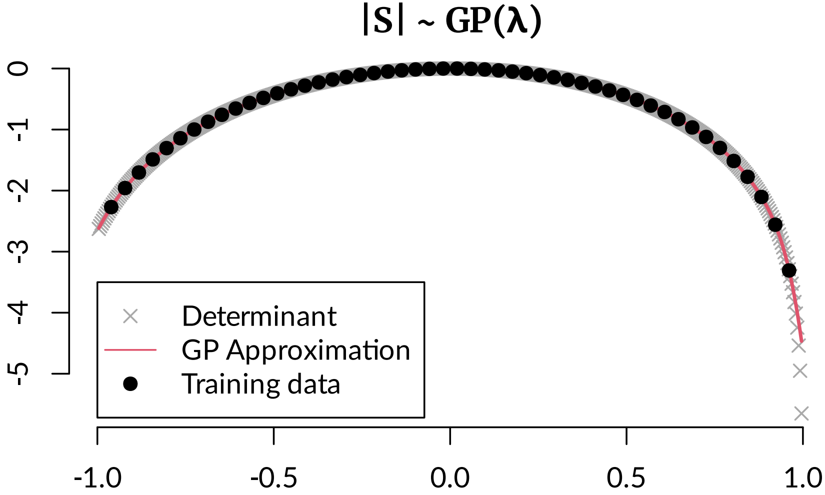


Figure 3: GP approximation to $|\mathbf{S}(\lambda)|$ using 50 training samples. Distances are between $N = 100$ locations with Uniform random coordinates.

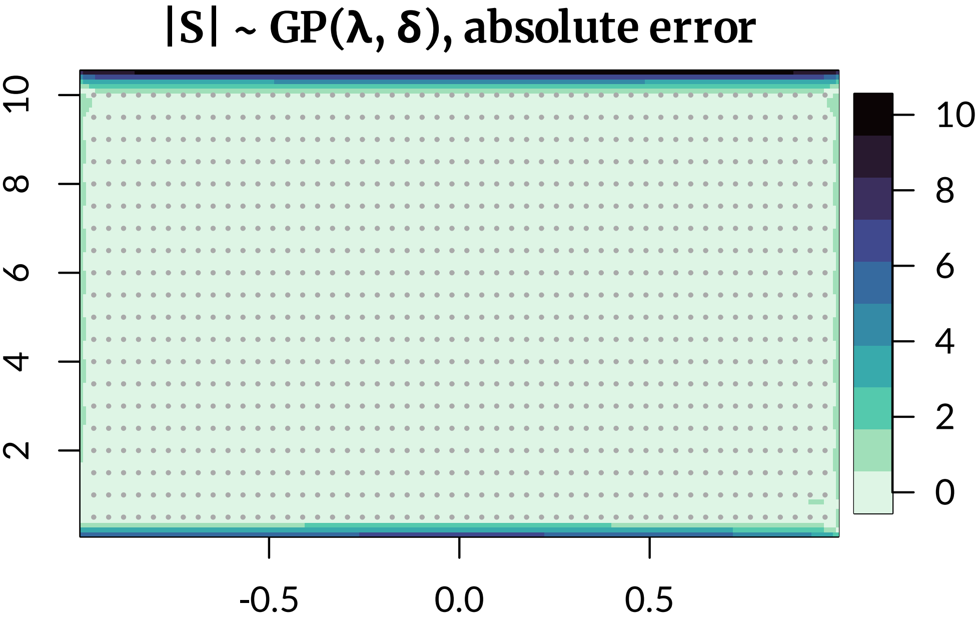
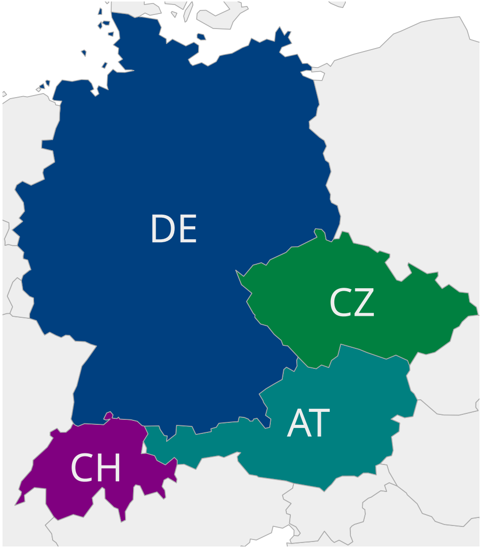


Figure 4: GP approximation to $|\mathbf{S}(\lambda, \delta)|$ using 50×20 training samples.

Imposing structure — Links Widely Shut

One way to reduce the dimensionality is by **constraining links** to **only occur within groups**.[a] ‣ Go back

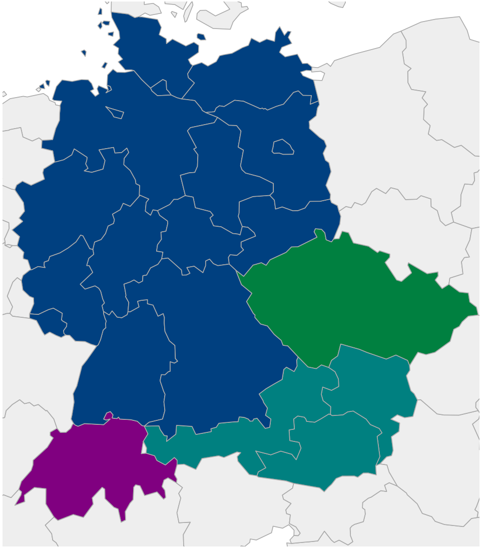


[a] This is essentially the approach of Lewbel et al., 2023.

Imposing structure — Links Widely Shut

One way to reduce the dimensionality is by **constraining links** to **only occur within groups.**[a] ▸ Go back

Country	NUTS 1
Austria	3
Czechia	1
South **Germany**[b]	2 (16)
Switzerland	1
Total ($N^2 - N$)	42
Grouped ($\sum_i N_i^2 - N_i$)	**10**



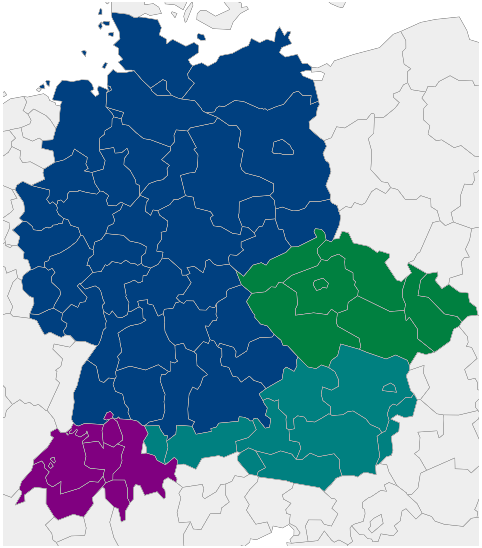
[a] This is essentially the approach of Lewbel et al., 2023.

[b] Thanks for nothing, re-unification.

Imposing structure — Links Widely Shut

One way to reduce the dimensionality is by **constraining links** to **only occur within groups**.[a] ‣ Go back

Country	NUTS 1	NUTS 2
Austria	3	9
Czechia	1	8
South Germany[b]	2 (16)	11 (38)
Switzerland	1	7
Total ($N^2 - N$)	42	1190
Grouped ($\sum_i N_i^2 - N_i$)	**10**	**280**



[a] This is essentially the approach of Lewbel et al., 2023.
[b] Thanks for nothing, re-unification.

Simulation results — US centroids

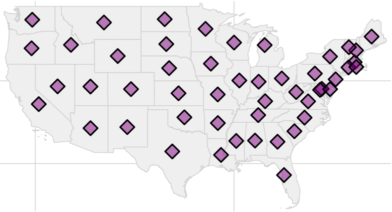
The second network is determined as

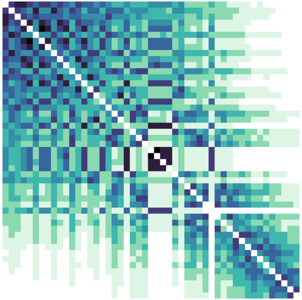
$$g_{ij} = \exp\left\{-\delta \times \mathrm{d}\left(\mathbf{p}_i, \mathbf{p}_j\right)\right\},$$

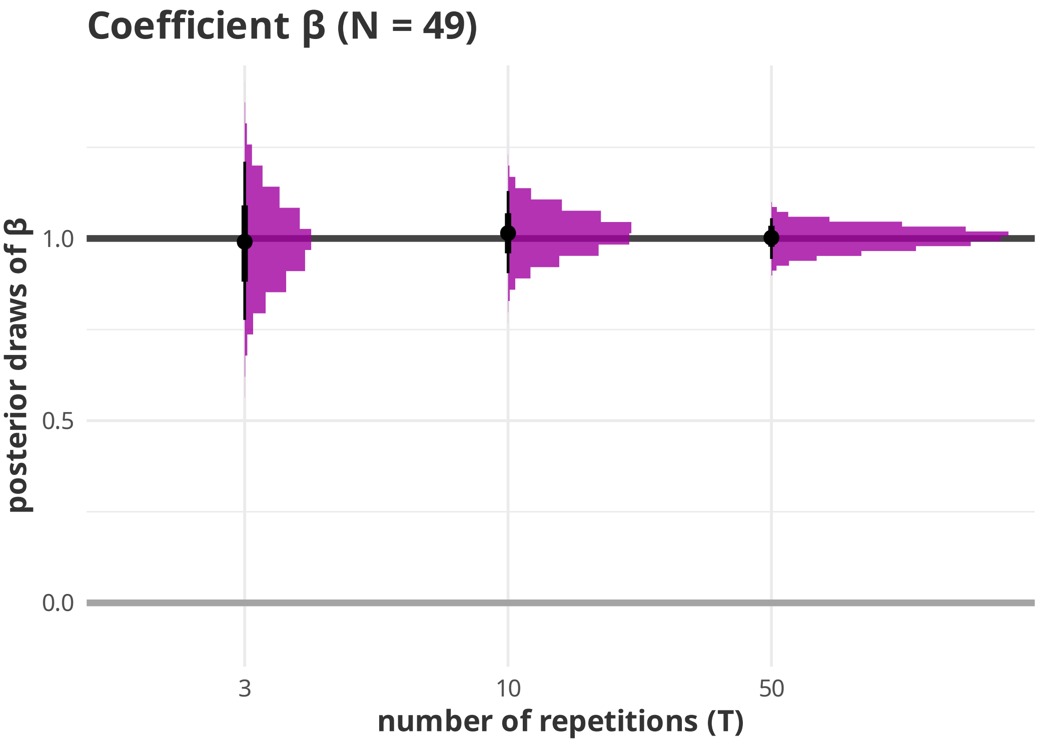
where $\mathbf{p}_i$ are **centroids**, and

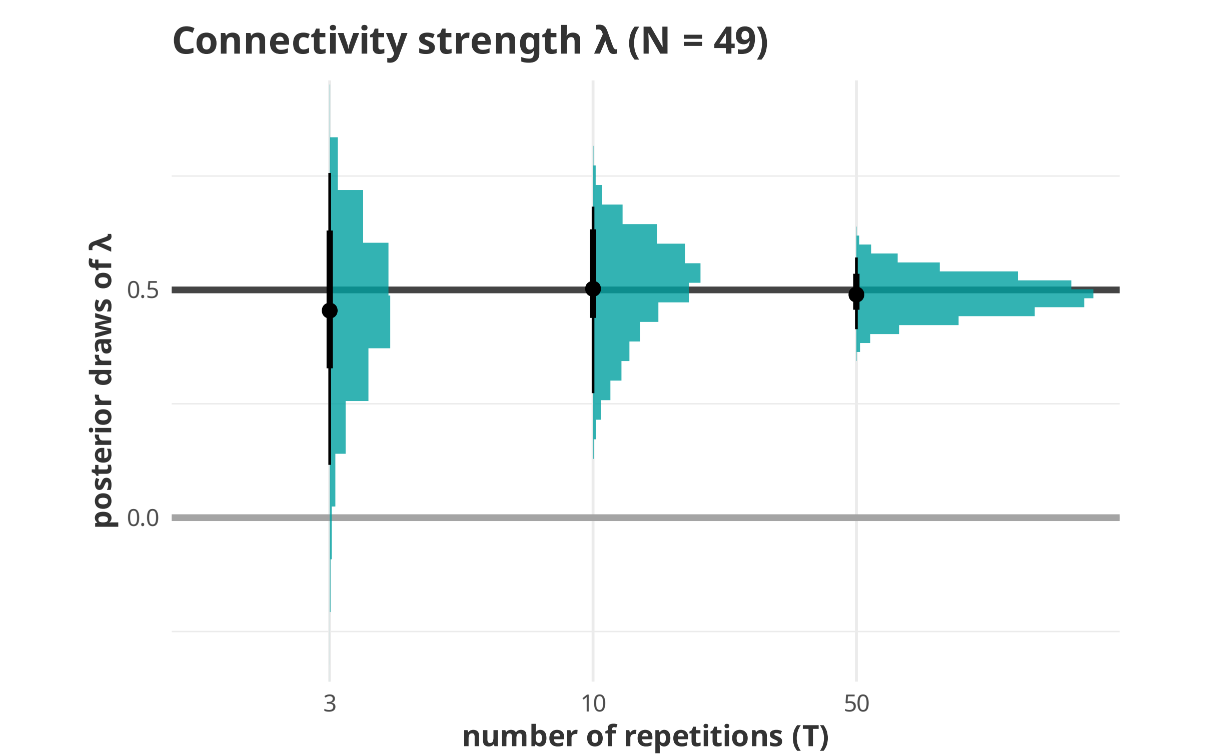
- the network is rather **dense**,
- I initialize the sampler at **random locations** to illustrate convergence.

We'll have a look at **posteriors of λ** for multiple simulations after a **short burn-in** of 1,000 draws. Go back






Coefficient β (N = 49)
posterior draws of β
1.0
0.5
0.0
3
10
50
number of repetitions (T)


Connectivity strength λ (N = 49)
posterior draws of λ
0.5
0.0
3
10
50
number of repetitions (T)

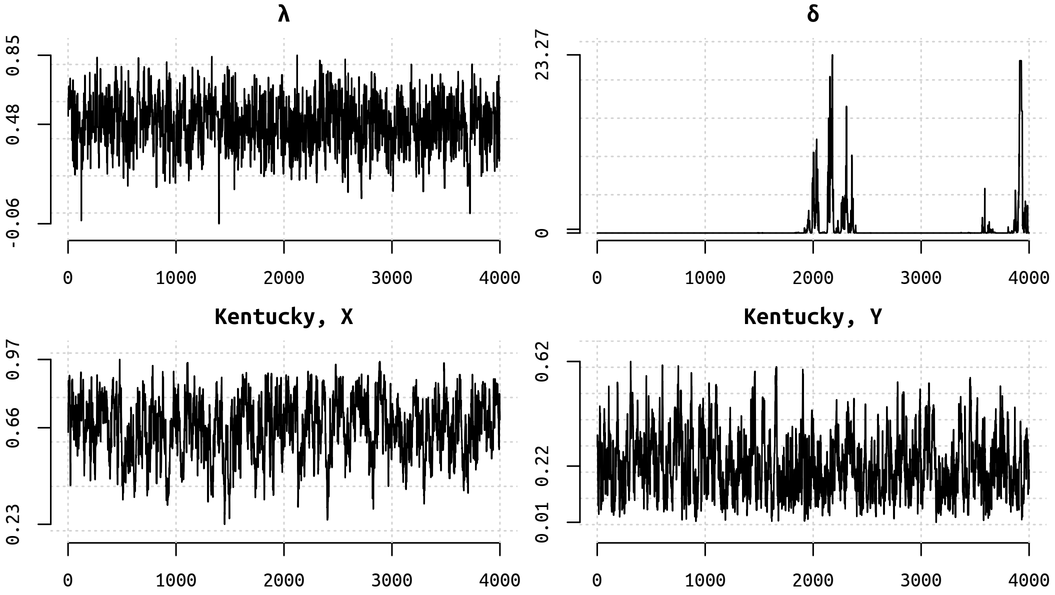


Figure 5: Traceplots for the posterior draws of λ, δ (note the poor mixing), and the (scaled) coordinates of Kentucky based on $T = 3$ network observations.

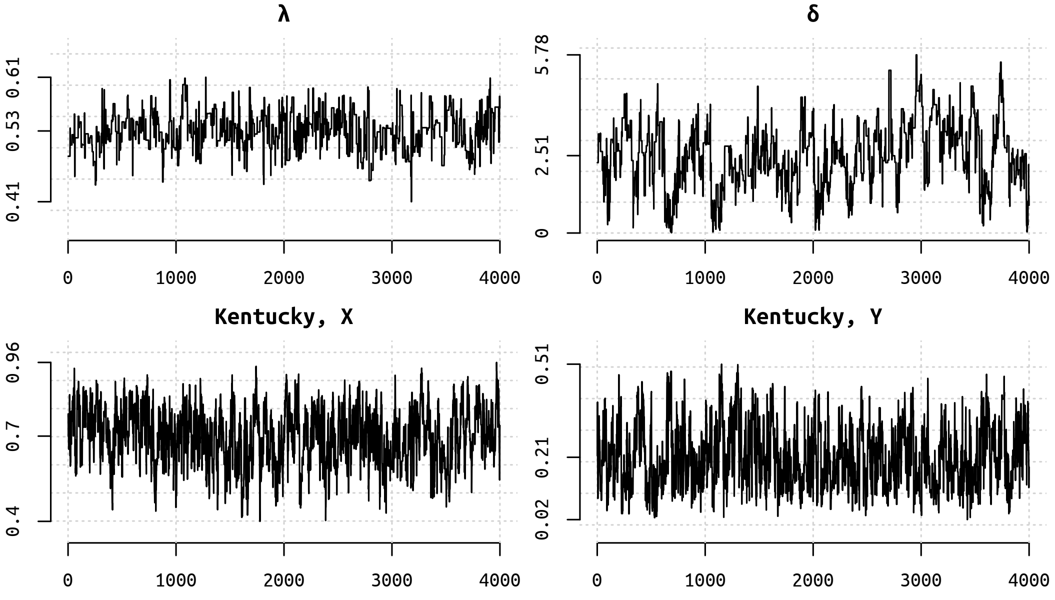


Figure 6: Traceplots for the posterior draws based on $T = 50$ network observations. Note the improved mixing behavior of δ. Go back