Spatial embeddings for cattle trade networks

Patrick Hoscheit, Étienne Pinsard, Anthony Muraro, Gaël Beaunée, Estelle Kuhn, Michele Salvi, Elisabeta Vergu

Spatial embeddings for cattle trade networks

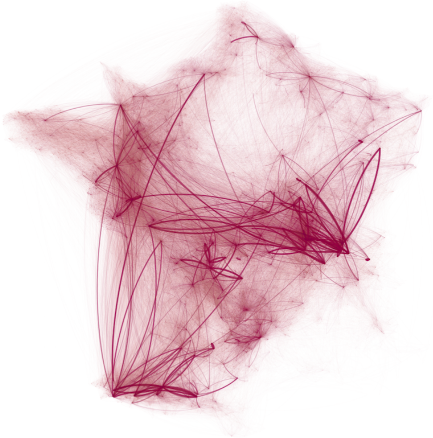
Patrick Hoscheit, Étienne Pinsard, Anthony Muraro, Gaël Beaunée,
Estelle Kuhn, Michele Salvi, Elisabeta Vergu

Modelling in Animal Health conference – ModAH^2 2021

French cattle trade network

- $\sim 250,000$ nodes (holdings)
- $\sim 18,000,000$ animals at any given time
- $\sim 5,000,000$ time-stamped edges (exchanges) per year

Movements between holdings logged on a daily basis since 2005 (Base de Données Nationale d'Identification, BDNI)



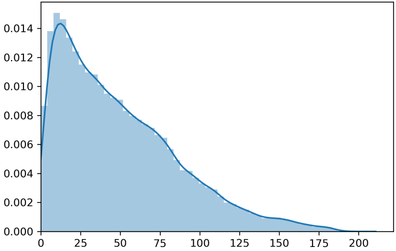
(Figures by Gaël Beaunée)

Spatial structure

Cattle trade network in Brittany, 2015 :

- 17,245 holdings spread over ~27,000 sq km
- Most movements on short scale (<100 km)
- Long edges mostly between high-volume nodes (markets/assembly centres)

Goal : formulate a generic model for cattle trade taking spatial and commercial inhomogeneity into account



Scale-free percolation

Given a set of vertices $V \subset \mathbf{R}^2$, draw ***weights*** $(W_x,\ x \in V)$ independently, according to a Pareto distribution with parameter $\tau - 1 > 0$:

$$\mathbb{P}(W \in dx) = \frac{\tau - 1}{x^\tau} dx.$$

Conditionally on the weights, connect vertices $x, y \in V$ with probability :

$$p_{x,y} = \mathbb{P}(x \leftrightarrow y) = 1 - \exp\left(-\lambda \frac{W_x W_y}{\|x - y\|^\alpha}\right).$$

- First introduced by [Deijfen et al. 2013] for $V = \mathbf{Z}^d$
- Studied by [Deprez and Wüthrich 2019] and [Dalmau and Salvi 2021] when V is Poisson-distributed on $\mathbf{R}^d$

Properties of SFP

If G is distributed as SFP(τ, λ, α) on the vertex set V :

- If $\alpha > 2$ and $\gamma = \alpha(\tau - 1)/2 > 1$, all vertices have finite degree a.s.
- G has scale-free degree distribution with exponent γ :

  $$\mathbb{P}(D_x > s) = \ell(s)s^{-\gamma},$$

  where ℓ is slowly varying at infinity.
- If $\gamma \in (1, 2)$, the graph distance between two points x, y scales like $\log\log \|x - y\|$ (***ultra-small world***)
- If $\gamma \geq 2$ and $\alpha \in (2, 4)$, graph distance scales like $\log \|x - y\|$ (***small world***)
- G has positive clustering (if $x \leftrightarrow y$ and $x \leftrightarrow z$, chances are that $y \leftrightarrow z$ also)

Simulation of SFP graphs

$$p_{x,y} = 1 - \exp\left(-\lambda \frac{W_x W_y}{\|x - y\|^\alpha}\right).$$

Naive sampling strategy inefficient for large networks :

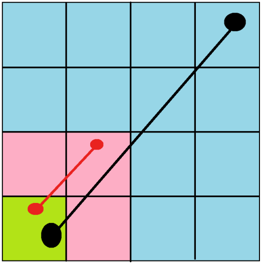
- Requires $\mathcal{O}(|V|^2)$ computations of $p_{x,y} = \mathbb{P}(x \leftrightarrow y)$
- For typical parameter values, only a small fraction of edges present (sparsity)
- Empirically, simulation time ∼20s for $|V| \sim 1000$
- Makes simulation-based inference methods impossible to use

Simulation of SFP graphs

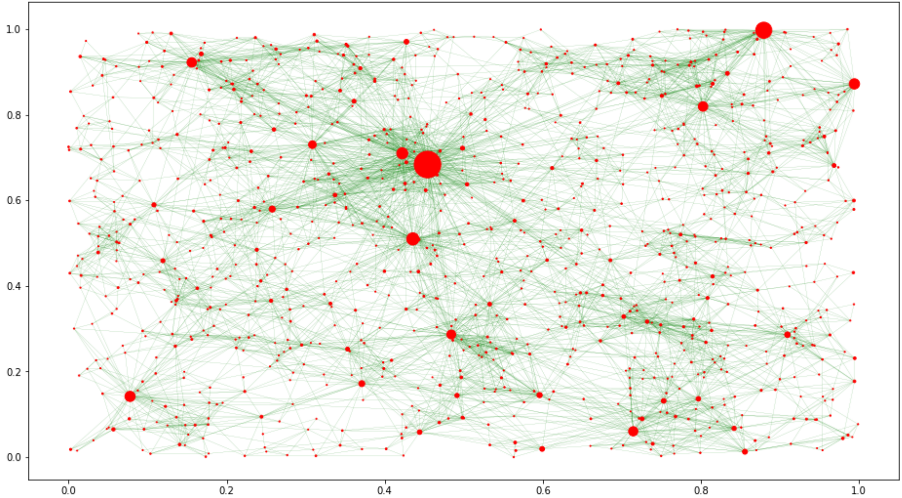
Improved simulation scheme, adapted from [Bringmann et al. 2018] :

- Decompose space into cells
- For (x, y) in adjacent cells, sample naively : $\mathcal{B}(p_{x,y})$
- For distant cells A, B, compute upper bound $\bar{p}$ for the connection probabilities $(p_{x,y},\ x \in V_A,\ y \in V_B)$
- Select potential edges with probability $\bar{p}$
- Keep each potential edge (x, y) with probability $p_{x,y} / \bar{p}$

For well-chosen cell sizes, complexity $\mathcal{O}(|V|)$



Example of a simulation

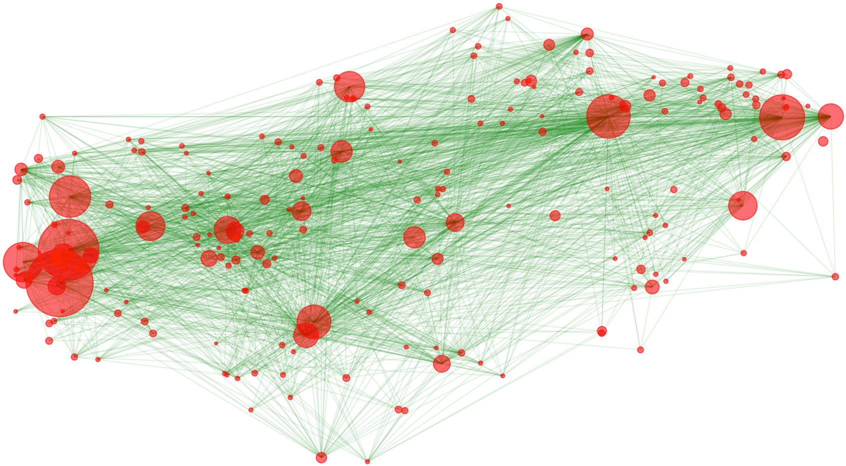


Model calibration

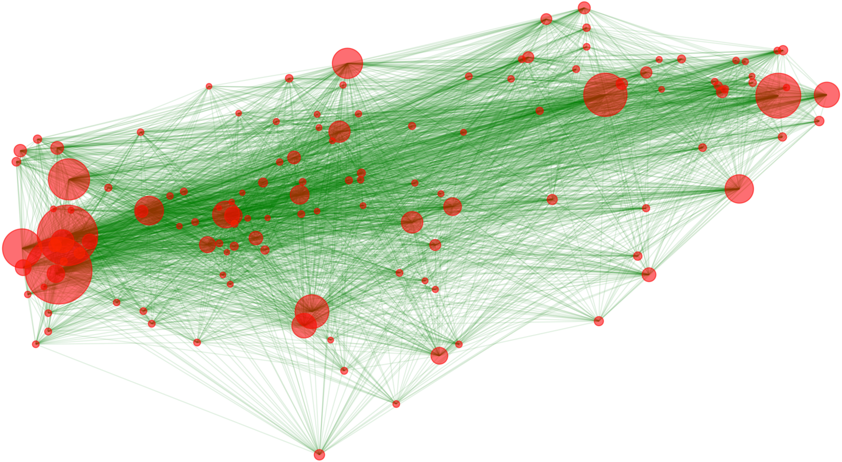
Use BDNI for node data :

- Locations (x_i, y_i) of holdings at commune level (1,208 in Brittany), randomised within commune boundaries
- Use ***trading volume*** V_i (number of animals entering and leaving a given holding over 1 year) as weights, which has power-law distribution with $\tau \sim 2.06$
- Best fit obtained for $\alpha = 1$ (away from $\alpha > 2$, $\gamma > 1$ range considered by Dalmau-Salvi)
    - Mean clustering 0.262 (true value 0.279)
    - Mean exponent of degree distribution 1.30 (true value 1.55)
    - Distance distribution qualitatively reproduced, but low variance

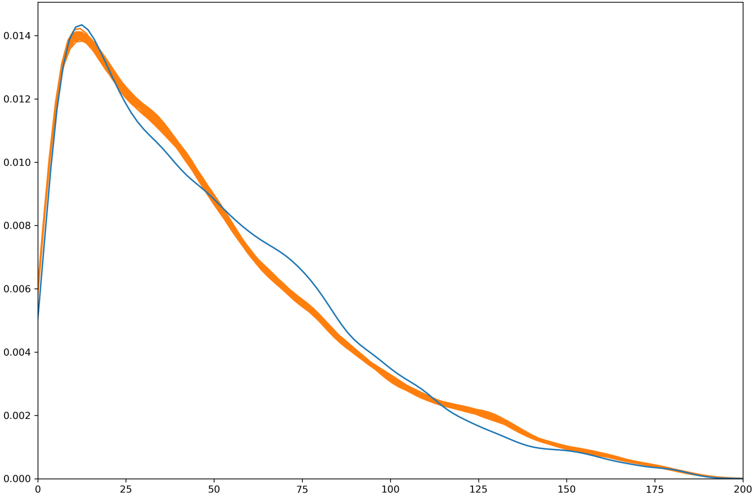
2015 observed network



Simulated network



Distance distribution



Parameter estimation

Hierarchical structure of the SFP model :

$$\mathbb{P}(W_x > s) = s^{-\tau}$$

$$\mathbb{P}(x \leftrightarrow y | W_x, W_y) = 1 - \exp\left(-\lambda \frac{W_x W_y}{\|x - y\|^{\alpha}}\right)$$

- Use HMM theory to infer latent parameter τ from observation of absence/presence of edges
- Maximum-likelihood estimation of τ, given α, λ, using SAEM-MCMC algorithm
- Estimation of α and λ computationally very intensive (use of ***minibatch*** technology [Kuhn et al. 2020])
- Joint inference out of reach at the moment

Conclusion

Scale-free percolation is a flexible model for spatially embedded trade networks :

- Model structure reflects inhomogeneity typical of these networks
- Key statistical features can be recovered
- Can be used for data analysis or forward simulation

Future work :

- Improve parameter estimation
- Incorporate high-resolution geospatial data to improve model fit at short scales
- Understand mechanistic drivers behind spatial structure
- Formulate a temporal version of the model

Thank you for your attention