Bayesian Financial Network Tomography

Reconstructing Networks from Partial Information

James Scott
and
Axel Gandy

Department of Mathematics
Imperial College London

STAT 300: Harvard, February 4, 2021

Outline

Table of Contents

Motivation

Economists & policy makers estimate contagion and systemic risk in economic/financial networks.

Contagion channels: Interbank lending, derivative exposures, equity cross-holdings.

Structure of links (interconnectedness) matters.

Statistical problem: *Only aggregated network data is available.* How to fill in missing data?

Need principled statistical reconstruction methods to recover networks from partial information.

Example Network

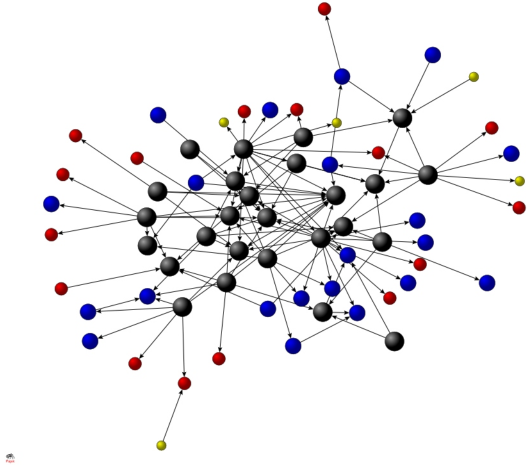

Figure: Topology of the Italian interbank money market (De Masi et al., 2006).

.

Table of Contents

Observe k statistics $T_i(G) = t_i$ from unknown graph G.

Assume G is drawn from an *ensemble* Ω according to P.

Choose P to maximise entropy whilst satisfying $t_1, ..., t_k$ in expectation.
Constrained Maximisation:

$$L(P) = -\mathbb{E}_P(\ln(P(G)) - \sum_{i=0}^{k} \lambda_i \left(\mathbb{E}_P(T_i(G)) - t_i\right),$$

with $T_0 = 1$ and $t_0 = 1$.

P has exponential family form:

$$P(G; \lambda) = \frac{1}{Z(\lambda)} \exp\left(-\sum_{i=1}^{k} \lambda_i T_i(G)\right). \tag{1}$$

Limitations of ERGs

ERG approach equivalent to:

1. Assuming parametric family (1)
2. Estimating parameters λ_i using MLE (based on one sample)
3. Simulate networks from $P(\mathrm{d}G; \hat{\lambda})$

Fails to *condition* on observed data $t_1, ..., t_k$. (Reasonable to expect inference to place zero measure on graphs not satisfying this data).

Most methods assign deterministic weights to edges once topology has been constructed. (Serious implications for systemic risk assessment).

See Park and Newman (2004); Squartini et al. (2018) for a review.

Table of Contents

Interbank Setting

Banks labelled $[n]$. Infer random matrix $X \in [0, \infty)^{n \times n}$.

X_{ij} represents institution i's liabilty to j (i.e. weight of edge $i \to j$).

We observe

1. total assets and liabilities $r(X)$ and $c(X)$ of each bank.
2. subset of edges $\{X_{ij} : (i, j) \in \mathcal{F}\}$ where $\mathcal{F} \subseteq [n] \times [n]$

Loops are not permitted, i.e. $X_{ii} = 0$ for all $i \in [n]$.

Goal: Construct prior distribution P, and estimate expectations of some integrable function h conditional on $T := (r, c, X_{\mathcal{F}})$.

The (Basic) Prior Model

Generate topology (adjacency matrix (A_{ij})) before attaching weights

$$A_{ij} \sim \text{Bernoulli}(p_{ij}),$$
$$X_{ij} | \{A_{ij} = 1\} \sim \text{Exponential}(\lambda_{ij}).$$

Interpretation: Generalised Erdős-Rényi Model.

Parameters $p = (p_{ij})$ and $\lambda = (\lambda_{ij})$.

Homogeneous setting: $p_{ij} = p_0 \mathbb{I}(i \neq j)$ and $\lambda_{ij} = \lambda_0$.

Hierarchical Models:

- p random gives control over degree distributions (power laws etc..).
- λ random gives control over weight distribution.

State space $\mathcal{X}_t = \{x \in [0, \infty)^{n \times n} : T(x) = t\}$ is a convex polytope.

MCMC algorithms exist for such problems, but not directly applicable due to the posterior distribution.

Diaconis and Gangolli (1995) introduced simple Gibbs sampler for contingency tables. Delineates 2×2 sub-matrix and updates take the form

$$\begin{matrix} +\Delta & -\Delta \\ -\Delta & +\Delta \end{matrix}$$

To construct an irreducible chain, we *require more general updates* than 2×2.

A Block Gibbs Sampler

Fix $k \in \{2, ..., n\}$ and randomly sample mutually disjoint row indices $(i_1, ..., i_k)$ and column indices $(j_1, ..., j_k)$.

Update distribution of X_n along coordinates

$$z := \{(i_1, j_1), (i_1, j_2), (i_2, j_2), ..., (i_k, j_k), (i_k, j_1)\}$$

conditional on other values.

Distribution concentrates on a line $X_{n+1,z_i} = X_{n,z_i} + (-1)^{i+1}\Delta$ with $\Delta \in \mathbb{R}$.

Suffices to consider conditional distribution of Δ.

Conditional Updates

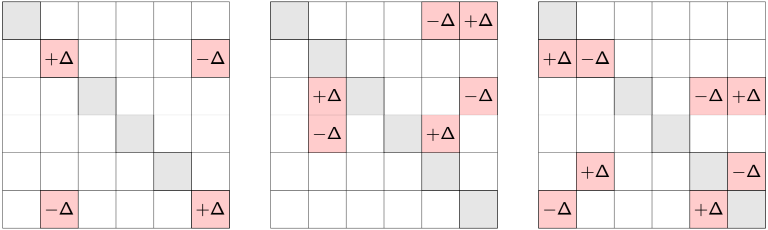

Figure: Example k-cycles of length 2 (left), 3 (center) and 4 (right) on a 6×6 matrix

Sampling Δ

The random variable Δ has interval range

$$\Delta \in [\Delta_l, \Delta_u] := \left[- \min_{i \text{ odd}} X_{z_i}, \min_{i \text{ even}} X_{z_i}\right].$$

It turns out $\mathbb{P}\{\Delta = 0\} = 1$ if

1. there is more than one zero entry in the cycle. i.e. $\#\{i : X_{z_i} = 0\} > 1$.
2. there is a known entry in the cycle. i.e. $z_i \in \mathcal{F}$ for some i.

Sampler will be prohibitively slow in large networks or with complex patterns of known edges.

State Dependent Mixing

Kernels $K := \{K_z : z \in \mathcal{Z}\}$ on a measurable space $(\mathcal{X}, \mathcal{B})$.

Define conditional distribution $w := \{w_x\}$ on $\mathcal{Z}$ given X_n.

Kernel Q defined by

$$Q(x, \mathrm{d}x') := \int_{\mathcal{Z}} K_z(x, \mathrm{d}x') w_x(\mathrm{d}z) \text{ for all } x \in \mathcal{X}.$$

Interpretation: Given X_n, sample $Z \sim w_{X_n}$ and $X_{n+1} \sim K_Z(X_n, dx')$.

Example (Random Scan Gibbs)

$K = \{\pi(dx_z | x_{-z}) : z \in [d]\}$ and $w_x(z) = w(z)$ for all x and z.

Example Decomposition

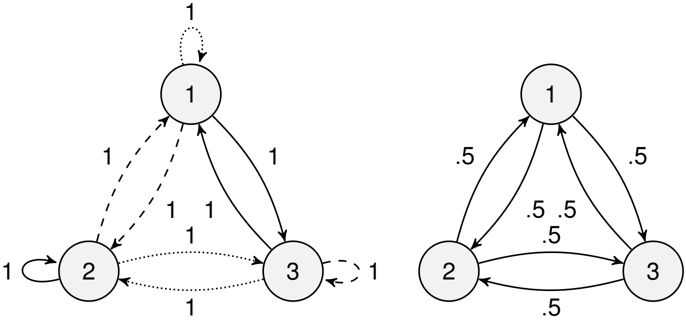

Figure: Kernels used as examples of a decomposition. Left: K_1 (dotted), K_2 (solid), K_3 (dashed). Right: Kernel Q

How to ensure Q has invariant distribution π?

Add kernel selection to the state space.

$Y_n := (Z_n, X_n)$ on $\mathcal{Z} \times \mathcal{X}$ with conditional transition probabilities

$$\mathbb{P}\{Y_{n+1} \in A_1 \times A_2 \mid Y_n = (z, x)\} = \int_{A_1} K_z(x, A_2) w_x(\mathrm{d}z),$$

for A_1 and A_2 measurable.

(Y_n) ergodic w.r.t. $w \otimes \pi \implies (X_n)$ ergodic w.r.t. π.

Algorithm

Suppose each K_z has invariant distribution $\pi(dx' | T_z(x))$ (assumed to exist) for some $T_z : \mathcal{X} \to E$ measurable.

Note: In our problem $T_z = X_{-z}$

It is easy to show that kernel L_z defined through

$$L_z(x, x') \propto w_{x'}(z) K_z(x, x')$$

has $w \otimes \pi$ as invariant distribution.

Algorithm 1 A Gibbs sampler for choice of scan order

Require: X_n
1: Sample $Z \sim w_{X_n}$
2: Sample $X_{n+1} \sim L_Z(X_n, \mathrm{d}x')$
return X_{n+1}

Scan Order and Mixing Time

Diaconis (2013):

> *"In the only cases where things can be proved, random scan and systematic scan have the same rates of convergence."*

Levin and Wilmer (2009) and Diaconis (2013) conjecture that systematic scan

- never mixes more than a *constant factor slower* than random scan
- never mixes more than a *logarithmic factor faster* than random scan.

He et al. (2016) disprove this conjecture with counterexamples.

Examples where state dependent mixing is *polynomial* faster than random scan.

Standard Hit and Run sampler is a special case:

- $\mathcal{X} \subseteq \mathbb{R}^d$.
- $\mathcal{Z}$ consists of all lines in $\mathbb{R}^d$
- w_x uniform on lines through x
- K_z has stationary distribution π restricted to line z.

Table of Contents

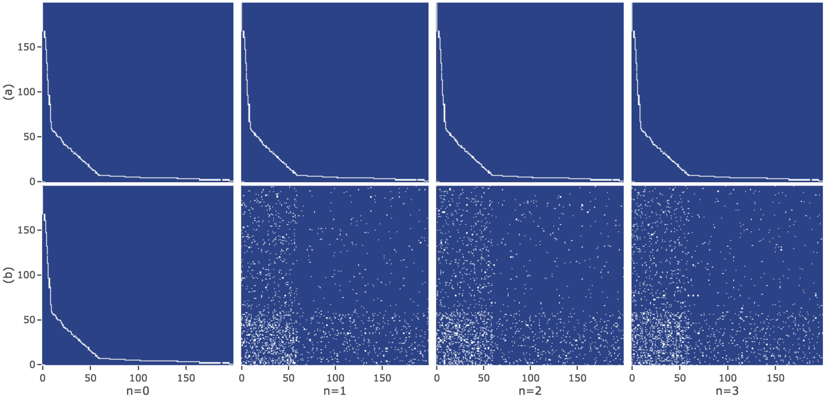

Figure: Adjacency matrices corresponding to samples from a size 200 network. Thinning chosen to time normalise samples. Panels (*a*) and (*b*) display consecutive samples using random scan and state-dependent mixing respectively. White entries correspond to edges. In (b) the mode is found rapidly whilst the topology in (a) is largely static between iterations.

Table of Contents

Discussion

- Developed a scalable Bayesian method for reconstruction of economic and financial networks.
- Affords greater flexibility (over ERGs) to incorporate prior information.
- Methods also useful for more general graph sampling problems, including exact tests for contingency tables (Scott and Gandy, 2018).
- Thank you for listening!

Table of Contents

References

De Masi, G., Iori, G., and Caldarelli, G. (2006), “Fitness model for the Italian interbank money market,” *Physical Review E*, 74(6), 066112.

Diaconis, P. (2013), “Some things we’ve learned (about Markov chain Monte Carlo),” *Bernoulli*, 19(4), 1294–1305.

Diaconis, P., and Gangolli, A. (1995), “Rectangular Arrays with Fixed Margins,” in *Discrete Probability and Algorithms*, eds. D. Aldous, P. Diaconis, J. Spencer, and Steele, Vol. 72 of *The IMA Volumes in Mathematics and its Applications* Springer New York, pp. 15–41.

He, B. D., De Sa, C. M., Mitliagkas, I., and Ré, C. (2016), Scan order in Gibbs sampling: Models in which it matters and bounds on how much,, in *Advances in neural information processing systems*, pp. 1–9.

Levin, and Wilmer, E. L. (2009), *Markov Chains and Mixing Time* American Mathematical Society.

Park, J., and Newman, M. E. (2004), “Statistical mechanics of networks,” *Physical Review E*, 70(6), 066117.

Scott, J., and Gandy, A. (2018), State-Dependent Kernel Selection for Conditional Sampling of Graphs,. arXiv:1809.06758 [stat.ME].

Squartini, T., Caldarelli, G., Cimini, G., Gabrielli, A., and Garlaschelli, D. (2018), “Reconstruction methods for networks: The case of economic and financial systems,” *Physics Reports*, .