Financial network reconstruction using Expectation Propagation

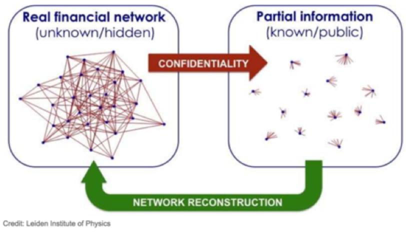
Mirko Pieropan

Supervisors: Andrea Pagnani and Alfredo Braunstein

Department of Applied Science and Technology, Politecnico di Torino, Italy

Motivation and background

The confidentiality of data poses a big challenge to the study of financial systems. Only partial information on interbank exchanges is usually available.



We would like to be able to reconstruct the whole network based on the limited number of observations at our disposal.

We can exploit assumptions on the sparseness of the system within a bayesian framework of the kind employed in (Braunstein, Muntoni and Pagnani, 2017).

Addressed problem

We wish to find solutions of the underdetermined linear system of equations:

$$\sum_{j=1}^{N_v} S_{ij}\, v_j = b_i, \qquad i = 1,\dots,M, \qquad N_v = N^2 - N, \qquad M = 2N.$$

S: observation matrix **v**: interbank fluxes, **b**: strengths

The solutions should satisfy a sparsity constraint. This is a compressed sensing problem, with density $\rho = K/N_v$ (K is the number of nonzero fluxes) and measurement rate $\alpha = M/N_v$.

State-of-the-art methods:

- Exponential random graphs (Park and Newman, 2004): maximize

$$S = -\sum_{G\in\mathcal{G}} P(G)\log P(G) \;\; s.t. \sum_{G\in\mathcal{G}} P(G)C_i(G) = C_i^*, \sum_{G\in\mathcal{G}} P(G) = 1.$$

  Formal solution:

$$P(G,\theta) = \frac{e^{-H(G,\boldsymbol{\theta})}}{Z(\boldsymbol{\theta})}, \qquad H(G,\boldsymbol{\theta}) = \sum_i \theta_i C_i(G).$$

  Find the the Lagrange multipliers θ_i s.t. $\langle C_i\rangle_{\theta^*} = C_i(G^*)\ \ \forall i$.
  - 'Bosonic' exponential random graphs: $C_i = b_i\ \forall i$. The model does not capture the sparsity of real networks.
  - Sparse generalization: constraints are b_i and the total number of links.
- Fitness model (De Masi et al., 2006):

$$x_i^{in}, x_i^{out} \text{ (fitnesses)}, \qquad x \sim \rho(x), \qquad p_{ij} = f(x_i^{out}, x_j^{in})$$

  estimate degrees and number of links via:

$$\sum_{j\neq i} f(x_i^{out}, x_j^{in}) = \langle k_i^{out}\rangle, \sum_{i\neq j} f(x_i^{out}, x_j^{in}) = \langle k_j^{in}\rangle, \sum_i \sum_{j\neq i} f(x_i^{out}, x_j^{in}) = \langle L\rangle$$

- Enhanced configuration model (Cimini et al., 2015):

$$f(x_i^{out}, x_j^{in}) = \frac{z x_i^{out} x_j^{in}}{1 + z x_i^{out} x_j^{in}}, \quad z \text{ free parameter;}$$

  identify fitnesses with strengths;
  estimate z by imposing constraint on number of links;
  use z to estimate the degrees as in the fitness model;
  use strengths and estimated degrees to compute ensemble probabilities p_{ij} and weights within the ERG model.

Materials and methods

Likelihood (it is a Boltzmann distribution):

$$P(\mathbf{b}|\mathbf{v}) = \left(\frac{\beta}{2\pi}\right)^{M/2} e^{-\beta E(\mathbf{v})}, \qquad E(\mathbf{v}) = \frac{1}{2}(\boldsymbol{S}\boldsymbol{v} - \boldsymbol{b})^T(\boldsymbol{S}\boldsymbol{v} - \boldsymbol{b})$$

Sparsity prior:

$$P(\mathbf{v}) = \prod_{k=1}^{N_v} \psi_k(v_k), \;\; \psi_k(v_k) = (1-\rho)\delta(v_k) + \frac{2\rho}{\sqrt{2\pi\lambda}}\, e^{-\frac{v_k^2}{2\lambda}}\, \mathbb{I}(v_k > 0)$$

Gaussian approximation of the posterior distribution:

$$Q(\mathbf{v}|\mathbf{b}) = \frac{1}{Z_Q} P(\mathbf{b}|\mathbf{v}) \prod_{k=1}^{N_v} N(v_k; a_k, d_k).$$

'Tilted' distribution:

$$Q^{(l)}(\mathbf{v}|\mathbf{b}) = \frac{1}{Z_{Q^{(l)}}} P(\mathbf{b}|\mathbf{v})\psi_l(v_l) \prod_{k\neq l} N(v_k; a_k, d_k).$$

Expectation Propagation (EP) algorithm:

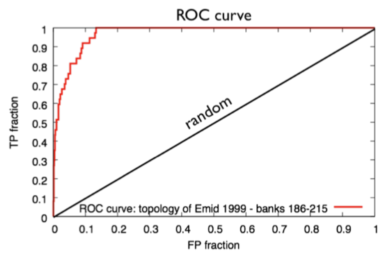
- until convergence: iteratively compute the means $\{a_1,\dots,a_{N_v}\}$ and variances $\{d_1,\dots,d_{N_v}\}$ such that $D_{KL}(Q^{(k)}||Q)$ is min w.r.t. a_k and d_k for all $k = 1,\dots,N_v$ ($\rightarrow$ moment matching);
- return means and variances of all v_k's w.r.t. the associated marginalized tilted distribution $Q^{(k)}(v_k|\mathbf{b})$.

The fixed points are stationary points of the EP free energy:

$$F_{EP} = (1 - N_v)F_Q + \sum_{k=1}^{N_v} F_{Q^{(k)}}, F_Q = -\log Z_Q, F_{Q^{(k)}} = -\log Z_{Q^{(k)}}$$

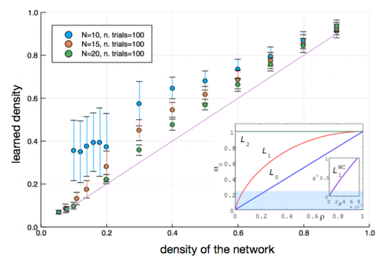
Preliminary results

What can we say about the (binary) topology of the network?
We looked at the e-MID 1999 dataset.
We are able to recover up to ~50% of the links.



ROC curve for a subnetwork having ρ=0.0425 and α=0.069.

The knowledge of the number of links can greatly improve the reconstruction, yet this information is seldom available.
By iteratively minimizing the EP free energy over the density parameter ρ of the priors, we can 'learn' the true number of links.



Learned density versus true density for several values of the system size.
Inset: space of solutions for the CS problem with L_0 and L_1 priors (for reference). The relevant region for the reconstruction of financial networks is the one highlighted in blue (adapted from Kabashima et al., 2009)

Future directions

Identify the convergence regions of EP in the (ρ, α) plane when using the sparsity priors relevant for the problem and attempt to build a possible-impossible reconstruction phase transition diagram.

SmartData@PoliTO infernet