# Thermodynamics of network model fitting with spectral entropies

Carlo Nicolini, Vladimir Vlasov, Angelo Bifone

carlo.nicolini@iit.it

Center for Neuroscience and Cognitive Systems
Istituto Italiano di Tecnologia

Paris, June 11, 2018

# Why models are important

**Complex networks models in neuroscience**

Generative network models are crucial in network neuroscience[1]

- Wiring rules or causal processes.
- Interest in the rules for building a network.
- A model can “compress” the description, highlighting regularities.

A new approach for evaluation of network models based on spectral properties.

[1] Betzel R., Generative Models for Network Neuroscience: Prospects and Promise, J. R. Soc. Interface 14: 20170623

# The spectral entropies method

PHYSICAL REVIEW X **6,** 041062 (2016)

**Spectral Entropies as Information-Theoretic Tools for Complex Network Comparison**

Manlio De Domenico[1,*] and Jacob Biamonte[2]

- Probability distributions encoded by density matrices ρ (unit trace, positive definite).
- Maximum uncertainty about a system with Hamiltonian $\mathbf{L}$ with the constraints:
- $\mathrm{Tr}\left[\rho\right] = 1$, $\langle \mathbf{L} \rangle = \mathrm{Tr}\left[\rho \mathbf{L}\right]$
- Quantum Gibbs-Boltzmann distribution:

$$\rho = \frac{e^{-\beta \mathbf{L}}}{\mathrm{Tr}\left[e^{-\beta \mathbf{L}}\right]}$$

- $\mathbf{L}$ is the graph Laplacian, semipositive definite symmetric matrix.

# Von Neumann entropy and relative entropy

## Von Neumann entropy

The Von Neumann entropy of the density matrix $\rho$ is:

$$S(\rho) = -\operatorname{Tr}[\rho \log \rho] = -\sum_{i=1}^{n} \lambda_i(\rho) \log \lambda_i(\rho)$$

It measures the departure of the system from a pure state.

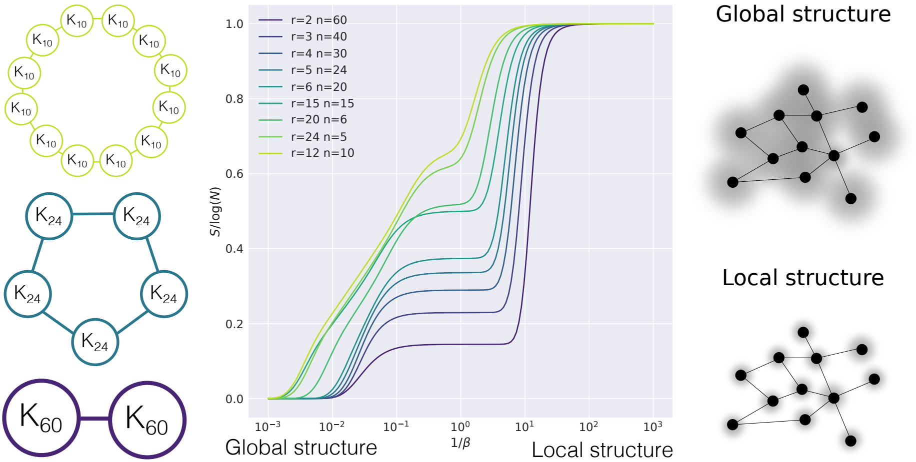
## Relative entropy

The relative entropy of the density matrices $\rho$ and $\sigma$:

$$S(\rho \| \sigma) = \operatorname{Tr}[\rho(\log \rho - \log \sigma)] \geq 0$$

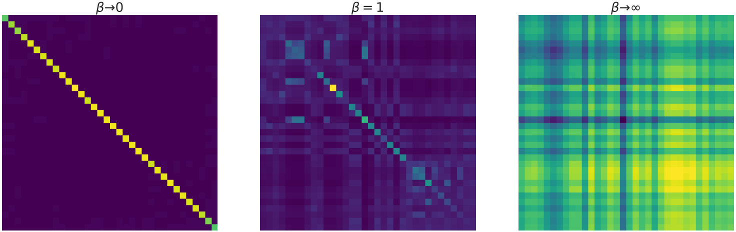
It measures the amount of information lost when $\sigma$ is used instead of $\rho$.

# Example of Von Neumann entropy in networks



# Motivation

$$S(\rho) = -\operatorname{Tr}[\rho \log \rho] = -\operatorname{Tr}\left[\frac{e^{-\beta \mathbf{L}}}{\operatorname{Tr}\left[e^{-\beta \mathbf{L}}\right]} \log\left(\frac{e^{-\beta \mathbf{L}}}{\operatorname{Tr}\left[e^{-\beta \mathbf{L}}\right]}\right)\right]$$




- There is a free parameter $\beta$: what is its role?
- Give a thermodynamic interpretation of relative entropy optimization.
- Provide a practical optimization method.

# Statistical thermodynamics link

With the thermal equilibrium density matrices, the expression of relative entropy becomes:

$$S(\rho\|\sigma) = \beta\left[\left(F_\rho - F_\sigma\right) - \left(\langle\mathbf{L}_\rho\rangle_\rho - \langle\mathbf{L}_\sigma\rangle_\rho\right)\right] \geq 0.$$

where:

- $F_\rho = -\beta^{-1}\log Z_\rho$ is the free energy.
- $\langle\mathbf{L}\rangle_\rho = \mathrm{Tr}\left[\rho\mathbf{L}\right]$ is the expected "energy".

**Klein inequality and Gibbs' inequality**

The state of minimum relative entropy is found by minimization of the left-hand side of:

$$\langle\mathbf{L}_\sigma\rangle_\rho - F_\sigma \geq \langle\mathbf{L}_\rho\rangle_\rho - F_\rho.$$

A "simple" receipt for model fitting within the spectral entropy framework.

# Optimization

Model optimization in this settings corresponds to finding the optimal parameters $\hat{\theta}$ such that:

$$\hat{\theta} = \underset{\theta}{\operatorname{argmin}} \quad \mathbb{E}_{\theta}[S(\rho \| \sigma(\theta)]$$

$$= \underset{\theta}{\operatorname{argmin}} \quad \operatorname{Tr}\left[\rho\left(\log\rho + \beta \underbrace{\mathbb{E}_{\theta}[\mathbf{L}(\theta)]}_{\text{easy}} + \mathbf{I} \underbrace{\mathbb{E}_{\theta}\left[\log Z(\theta)\right]}_{\text{hard to compute}}\right)\right]$$

- Knowledge of Laplacian spectra via random matrix theory.
- Monte Carlo sampling.
- Either hard to obtain, or slow to compute: we use an approximation.

$$\mathbb{E}_{\theta}[S(\rho \| \sigma(\theta)] \approx S(\rho \| \sigma(\mathbb{E}_{\theta}[\mathbf{L}]))$$

# Gradients and exponential random graph models

Two variants of the exponential random graph models:

- Erdos-Renyi
- Planted partition model (two blocks)

By setting gradients of relative entropy to zero:

$$\frac{\partial S(\rho \| \sigma(\mathbb{E}[\mathbf{L}]))}{\partial \theta} = \beta \operatorname{Tr}\left[ \left(\rho - \sigma(\mathbb{E}[\mathbf{L}])\right) \frac{\partial \mathbb{E}[\mathbf{L}](\theta)}{\partial \theta} \right],$$

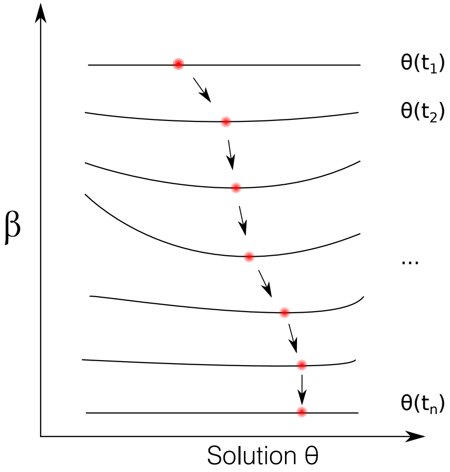
in the Erdos-Renyi model we find:

$$p^* = \hat{p} = \lim_{\beta \to 0} \frac{1}{n\beta} \log \left( \frac{\operatorname{Tr}[\mathbf{1}\rho]\,(n-1)}{(n - \operatorname{Tr}[\mathbf{1}\rho])} \right)$$

Similarly in the planted partition model, with two blocks and $p_{\text{in}}, p_{\text{out}}$ reconstruction in the limit $\beta \to 0$

# An optimization algorithm

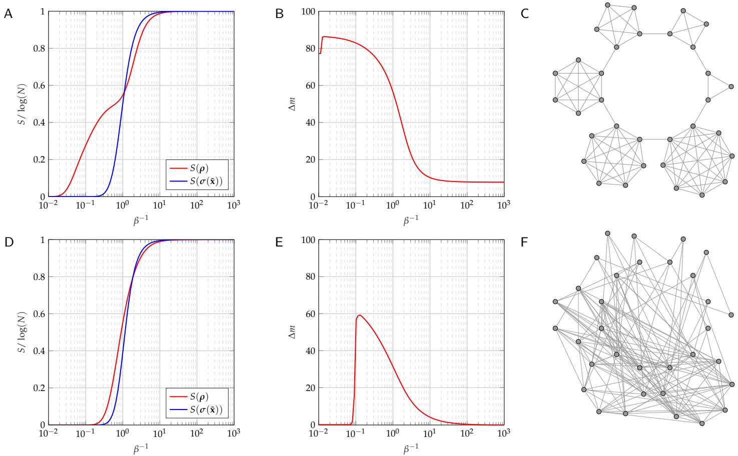
1. Start with some random solution $\theta(t_1)$ and low temperature $\beta \gg 1$.
2. Minimize $S(\rho\|\sigma)$ to get a new solution $\theta_1$.
3. Decrease $\beta$ by some small amount.
4. Return to step 2 while convergence is achieved.



# Undirected binary configuration model

Exponential random graph model with constrained degree sequence. Nodal hidden variables $x_i$:

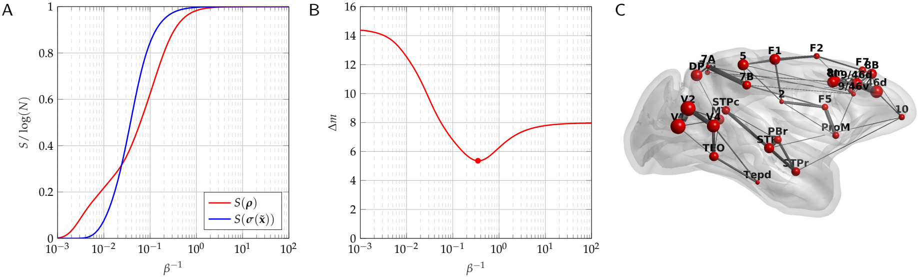
$$p_{ij} = \mathbb{E}[a_{ij}] = \frac{x_i x_j}{1 + x_i x_j}$$



# Macaque connectivity

Exponential distance rule:

$$\mathbb{E}[A_{ij}] = Ce^{-\ell d_{ij}}$$




$\ell \approx 0.15\ \mathrm{mm}^{-1}$ according to other methods.

# Conclusions

- A new framework the study of complex networks at different scales.
- An interpretation of the meaning of $\beta$.
- A practical implementation of gradient descent methods.

**Networkqit code-alpha release**

**Documentation**: `networkqit.github.io`
**Repository**: `bitbucket.org/carlonicolini/networkqit`
**Arxiv:** `https://arxiv.org/abs/1801.06009`

# Appendix: approximation via matrix concentration in random graphs

For matrices of iid variables, in the large $n$ and low sparsity limit all eigenvalues tend to their expected counterpart[2],[3]:

$$\Pr\left(|\lambda_i(\mathbf{L}) - \lambda_i(\mathbb{E}(\mathbf{L}))| \geq t\right) \leq \text{(some exponentially decaying function of t)}$$

For this reason we approximate $\lambda_i(\mathbf{L})$ with $\lambda_i(\mathbb{E}[\mathbf{L}])$, to get:

$$\log Z(\theta) = \log \sum_{i=1}^{n} e^{-\beta\lambda_i(\mathbf{L})} \approx \log \sum_{i=1}^{n} e^{-\beta\lambda_i(\mathbb{E}_\theta[\mathbf{L}])}$$

Hence:

$$\mathbb{E}_\theta[S(\rho\|\sigma(\theta)] \approx S(\rho\|\sigma(\mathbb{E}_\theta[\mathbf{L}(\theta)]))$$

[2] Cape et al. arXiv:1603.06100v1 (2017)

[3] Imbuzeiro Oliveira, arXiv:0911.0600 (2009)

# Validity of the approximation