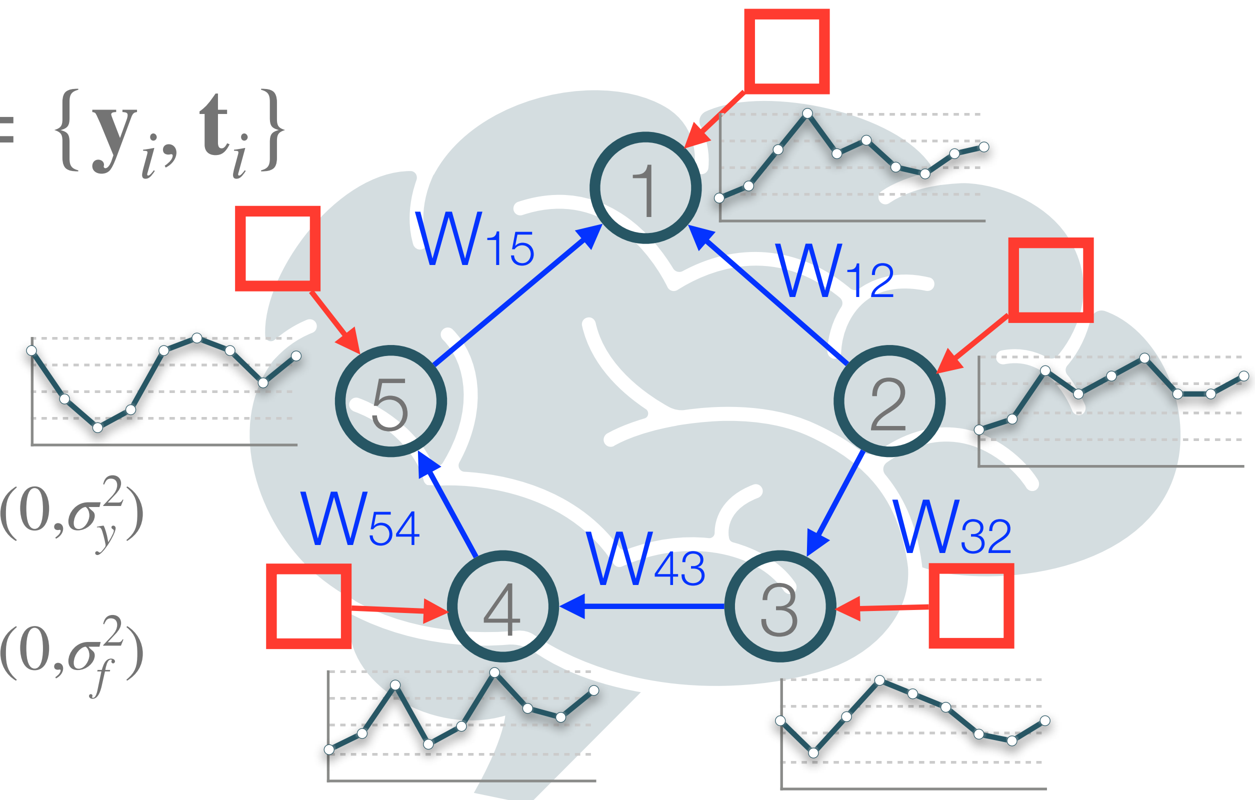
Variational Network Inference: Strong and Stable with Concrete Support

Amir Dezfouli, **Edwin V. Bonilla** and Richard Nock

Network Structure Discovery: A Flexible Approach

- N nodes, T observations: $\mathscr{D} = \{\mathbf{y}_i, \mathbf{t}_i\}$
- Goal: Learn network structure
  - Existence, directionality and strengths



- Model

$\epsilon_{it} \sim \text{Normal}(0, \sigma_y^2)$

$$y_i(t) = f_i(t) + \epsilon_{it}$$

$\xi_{jt} \sim \text{Normal}(0, \sigma_f^2)$

$$f_i(t) = z_i(t) + \sum_{j=1, j \neq i}^{N} A_{ij} W_{ij} \left[f_j(t) + \xi_{jt} \right]$$

Network-independent trend

$$z_i(t) \sim \text{GP}(0, \kappa(t, t'; \boldsymbol{\theta}))$$

Network parameters
$A_{ij} \in \{0, 1\}, \quad W_{ij} \in \mathbb{R}$

$$p(\mathbf{A}, \mathbf{W}) = \prod_{ij} p(A_{ij}) p(W_{ij})$$

$$p(A_{ij}) = \text{Bern}(\rho)$$

$$p(W_{ij}) = \text{Normal}(0, \sigma_w^2)$$

Inference Goal: Estimate $p(\mathbf{A}, \mathbf{W} \,|\, \mathscr{D})$

- Complications:
  - **f** defined cyclically
  - GPs notoriously unscalable $O(N^3T^3)$
  - Complicated marginal likelihood (**f** depends on **A**, **W**)

Trick 1: Derive "inverse" model

$$\mathbf{f}(t) = (\mathbf{I} - \mathbf{A} \odot \mathbf{W})^{-1}(\mathbf{z}(t) + \mathbf{A} \odot \mathbf{W}\boldsymbol{\xi}_t)$$

Trick 2: Marginalise **f** analytically

$$p(\mathbf{y} \,|\, \mathbf{A}, \mathbf{W}) = \text{Normal}(\mathbf{0}, \boldsymbol{\Sigma}_y)$$

$$\boldsymbol{\Sigma}_y = \mathbf{K}_f \otimes \mathbf{K}_t + \mathbf{K}_\sigma \otimes \mathbf{I}$$

Trick 3: Relate to Multi-task learning (MTL)

- MTL with product covariance (Bonilla et al, 2008; Rakitsch et al, 2013)
  - Nodes are "tasks"
- Sum of two Kronecker products
  - Covariances determined by **A**, **W**
- More efficient computation $O(N^3 + T^3)$

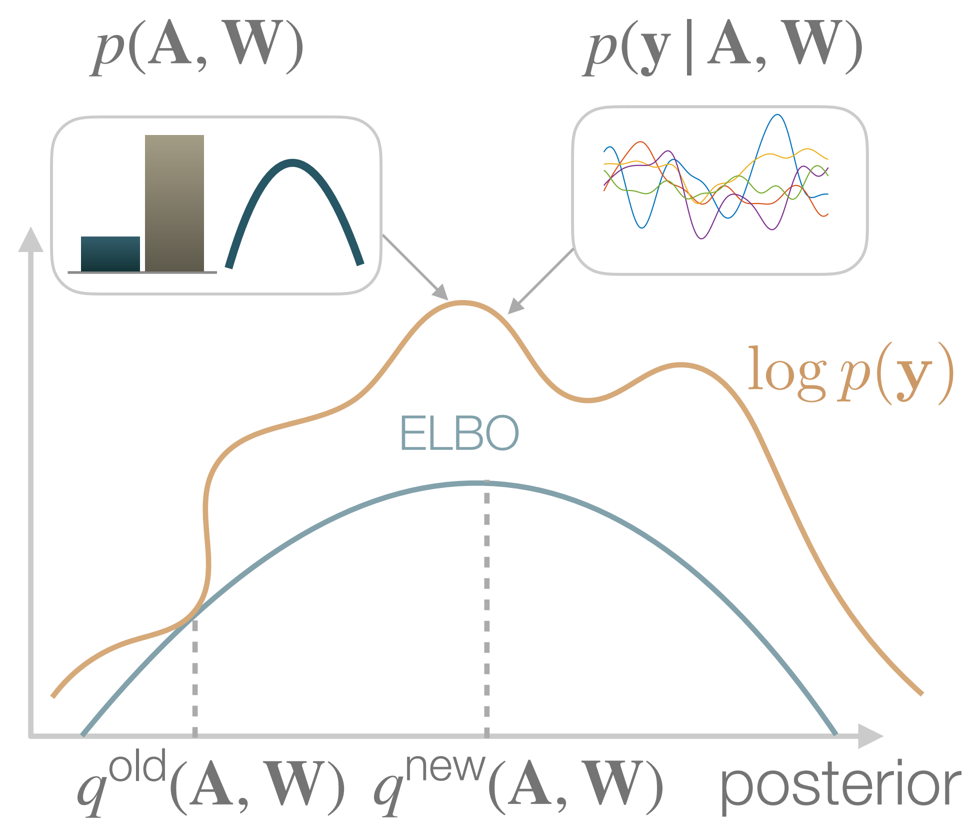
- How to deal with complex dependency on **A**, **W**?
  - *Modern variational inference*

Modern Variational Inference

$$\mathscr{L}_{\text{elbo}} = \mathscr{L}_{\text{kl}} + \mathscr{L}_{\text{ell}}$$

$$\mathscr{L}_{kl} = -\,\text{KL}(q(\mathbf{A},\mathbf{W})||p(\mathbf{A},\mathbf{W}))$$

$$\mathscr{L}_{ell} = \mathbb{E}_{q(\mathbf{A},\mathbf{W})} \log p(\mathbf{y}\,|\,\mathbf{A},\mathbf{W})$$



- Expectations using Monte Carlo
  - Re-parameterization trick
  - Cannot be applied to discrete rv

Trick 4: Concrete Distribution: $q(A_{ij}) = \text{Concrete}(\alpha_{ij}, \lambda_c)$

- α_{ij} are variational parameters
- Aka Gumble-Softmax trick, can sample and evaluate *log q(A$_{ij}$)*
- *It helps us get stability for free*

Theory: Numerical Stability

- Usually imposes the non-singularity of $\mathbf{I} - \mathbf{A} \odot \mathbf{W}$
  - Sometimes with additional constraints (boundedness of coordinates, eigenvalues)

Theorem 1: "We get stability for free"

For any $\lambda_c \geq 0, \alpha_{ij} \geq 0 (i \neq j)$, $\mathbf{I} - \mathbf{A} \odot \mathbf{W}$ *is non-singular with probability 1.*

Theorem 2:

For any $\lambda_c \geq 0, \alpha_{ij} \geq 0 (i \neq j)$, $\sigma_y^2 \geq 0$, $|\mathscr{L}_{ell}| \ll \infty$

Theory: Model Stability

- Bounds the signal's log likelihood as a function of external parameters

Theorem 3: Statistical "robustness"

If $W_{ij} \sim \text{Normal}(\mu_{ij}, \sigma_{ij}^2)$ *and* $A_{ij} \sim \text{Bern}(\rho_{ij})$, *then under a* ***condition*** *on the network signal, it holds with large probability:*

$$-\log p(\mathbf{y} \,|\, \mathbf{W}, \mathbf{A}) \in [g(\lambda_\circ, \mathbf{y}), g(\lambda_\bullet, \mathbf{y})] \;\; , \forall \mathbf{y}, \textit{ where}$$

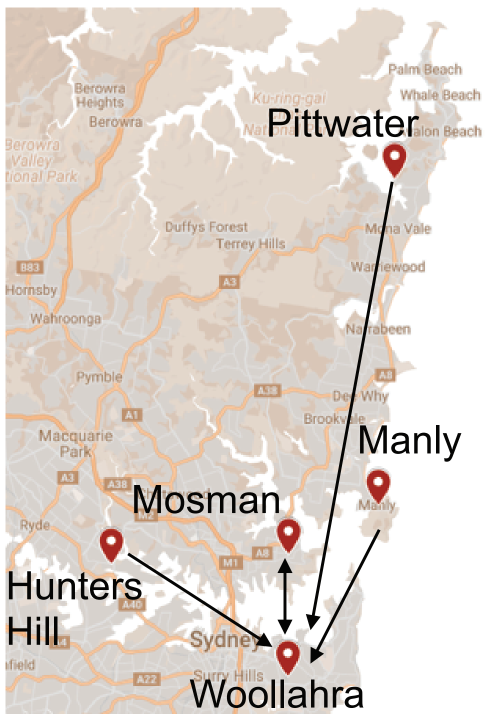
$$g(z, \mathbf{y}) = \theta(\log z + z\|\mathbf{y}\|_2^2),\ \lambda_\circ = \lambda^\downarrow(\mathbf{K}_t)/2 + \sigma_y^2,\ \lambda_\bullet = 2\lambda^\uparrow(\mathbf{K}_t) + \sigma_f^2 + \sigma_y^2$$

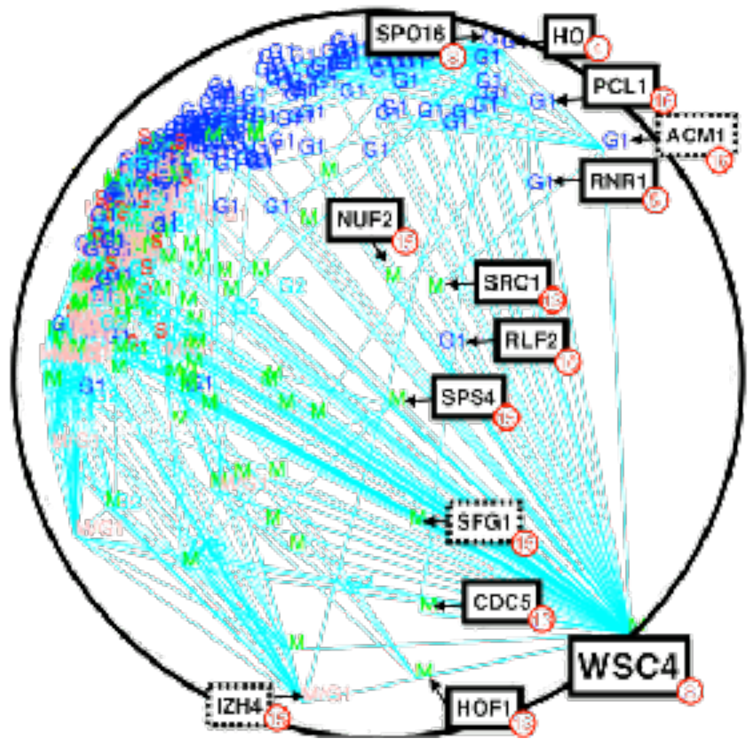
and ***condition*** *appears in various forms in previous work.*

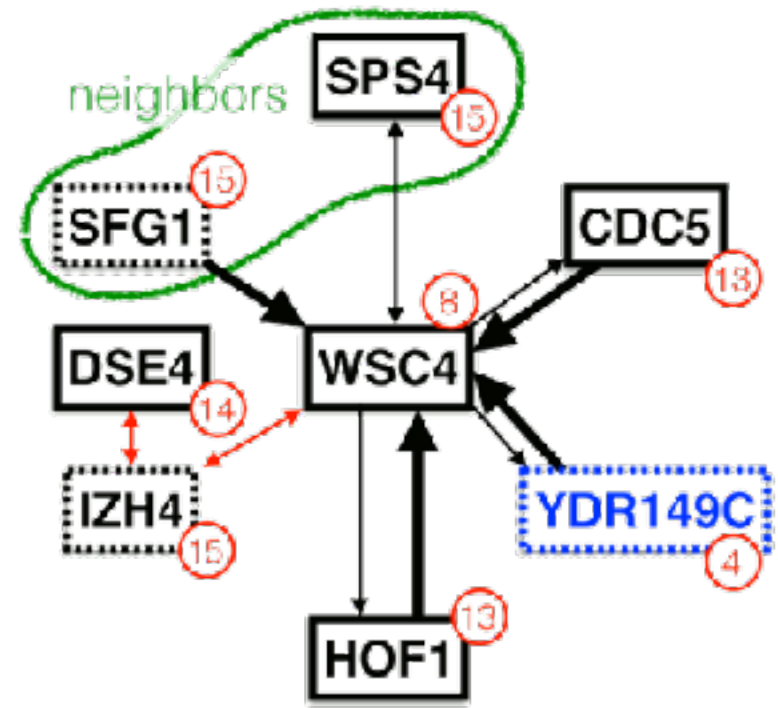
- Important practical consequences

Experiments and Conclusions

- Sydney property prices, brain fMRI, yeast genome







- Bayesian approach for network structure discovery
  - Efficient inference
- stability for “free”, robustness and easy estimation