Higher-order interactions in statistical physics, machine learning (and biomedicine)

Ava Khamseh
Schools of Informatics and Physics
Institute of Genetics and Cancer

28 Sep 2021

Collaborators

School of Physics

Guido Cossu

Luigi Del Debbio

Tommaso Giani

Michael Wilson

School of Mathematics

Sjoerd Beentjes

Institute of Genetics & Cancer

Chris Ponting

Abel Jansma

UC Berkeley

Mark van der Laan

Forward vs inverse problems

Forward problem (Statistical Physics): The goal is to provide a macroscopic description of Nature by deriving observable quantities from underlying laws.

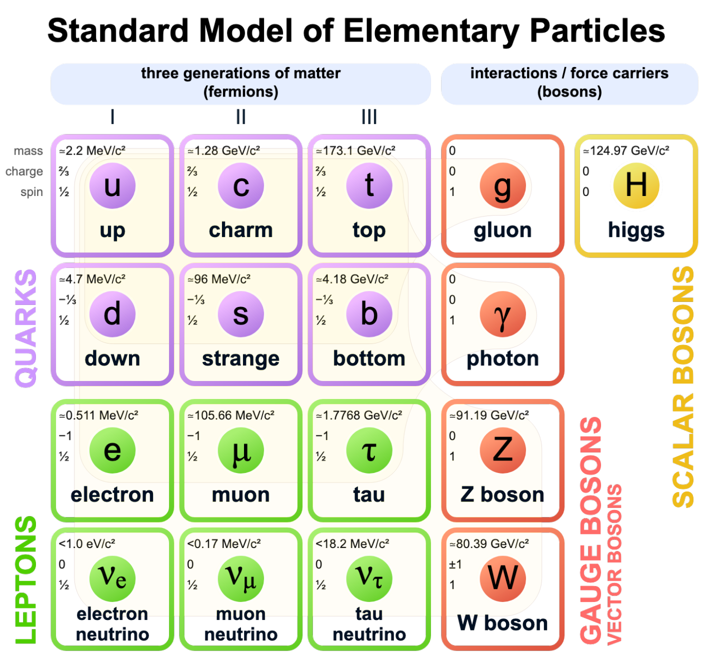
- Ising model forward problem: Obtain observables such as magnetisation, energy, correlations, given the Hamiltonian and its parameters

Inverse problem: Starting point are observations (data), the goal is to infer microscopic properties of the system

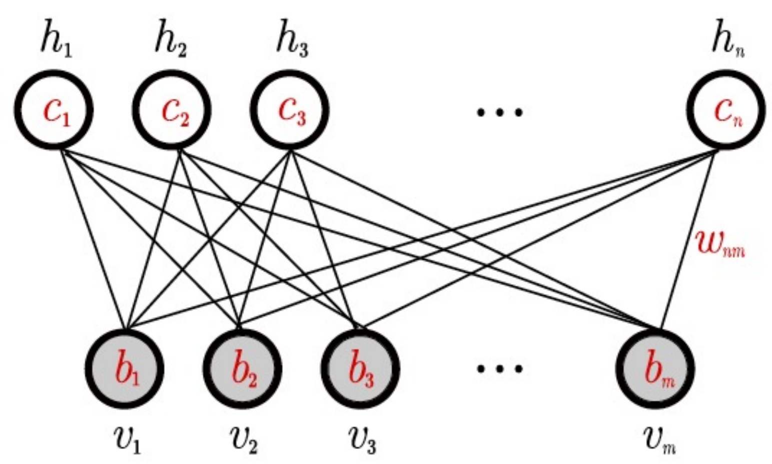
- Estimate Ising interactions directly from data

Interaction in science

Elementary Particles

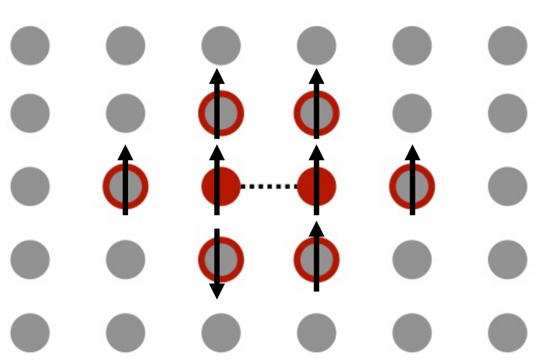


Nodes in a neural network

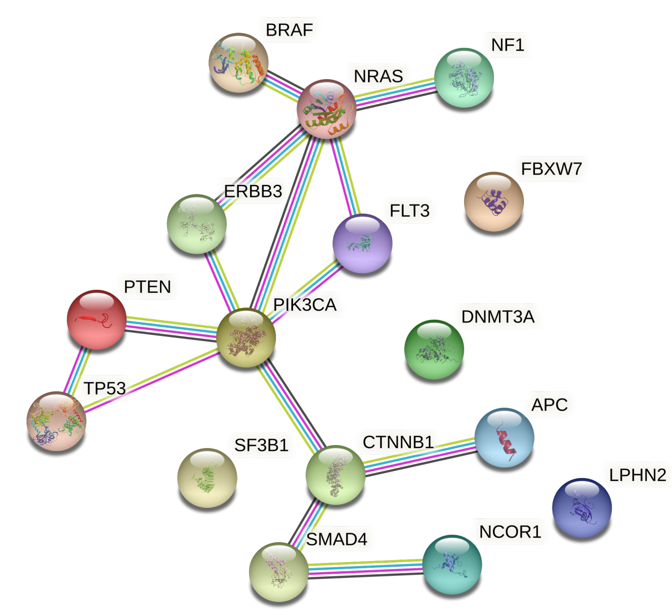


Interactions

Statistical Physics



Biomedicine

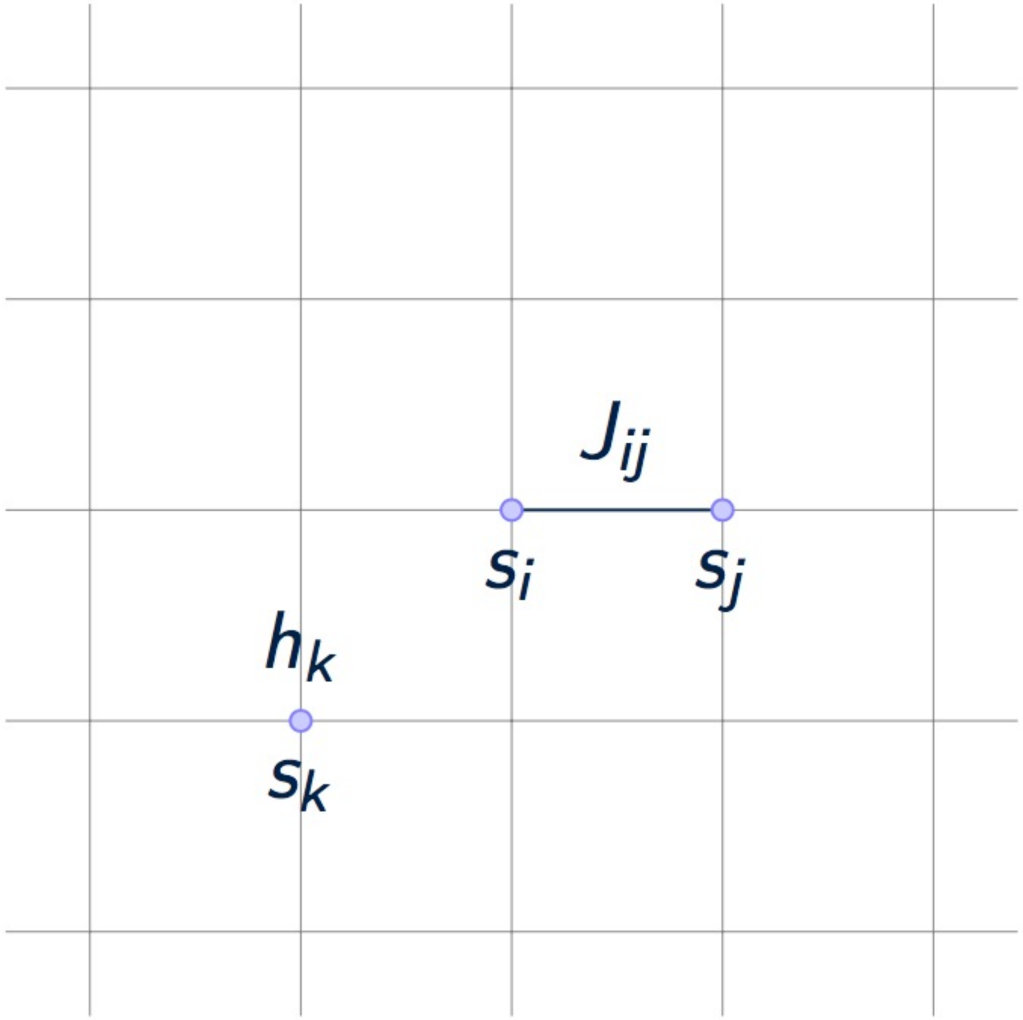


Interactions:
The Ising Model & RBM

Ising model

$$p_D(s) = \frac{1}{Z(J,h)} e^{-H_{J,h}(s)}$$

$$H_{J,h} = -\sum_{i,j} J_{ij} s_i s_j - \sum_i h_i s_i \quad , \quad Z(J,h) = \sum_s e^{-H_{J,h}(s)}$$



Onsager 1944

MC simulation of the 2D Ising model at various temperatures

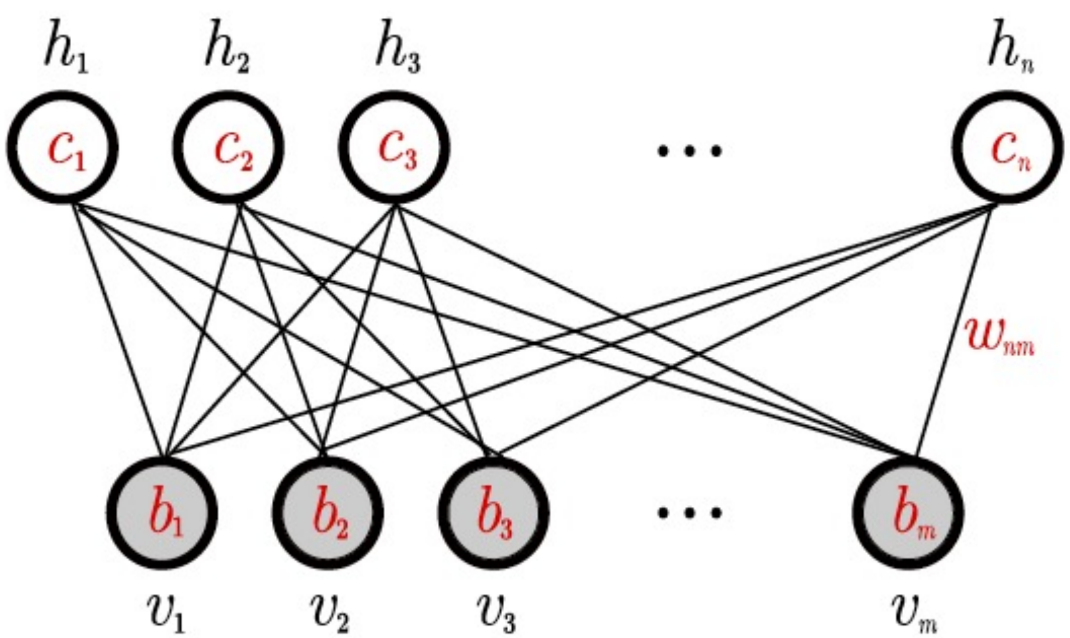
generate a sample
$D = \{s^1, s^2, \ldots\}$, $N_D \sim 10^5$

Machine Learning: Restricted Boltzmann Machine

Restricted Boltzmann Machine

$$E_\theta(\mathbf{v}, \mathbf{h}) = -\sum_{i=1}^{n}\sum_{j=1}^{m} w_{ij} h_i v_j - \sum_{i=1}^{n} c_i h_i - \sum_{j=1}^{m} b_j v_j$$

$$\boxed{p_{\text{RBM}}(\mathbf{v}, \mathbf{h}|\theta) = \frac{1}{Z_{\text{RBM}}} e^{-E_\theta(\mathbf{v},\mathbf{h})}}$$



$$D_{\text{KL}}\left(q_{\text{data}}(\mathbf{v})||p_\theta(\mathbf{v})\right) = \sum_{\mathbf{v}} q_{\text{data}}(\mathbf{v}) \log\left(\frac{q_{\text{data}}(\mathbf{v})}{p(\mathbf{v})}\right)$$

$$= \sum_{\mathbf{v}} \Big(q_{\text{data}}(\mathbf{v}) \log\left(q_{\text{data}}(\mathbf{v})\right) - q_{\text{data}}(\mathbf{v}) \log\left(p_\theta(\mathbf{v})\right) \Big)$$

Max likelihood $\Longleftrightarrow$ min KL divergence

Contrastive Divergence

$$\frac{\partial \log \mathcal{L}(\theta|\mathbf{v})}{\partial w_{ij}} = p(h_i = 1|\mathbf{v})v_j - \langle p(h_i = 1|\mathbf{v}')v_j \rangle_{p(\mathbf{v}')} \, .$$

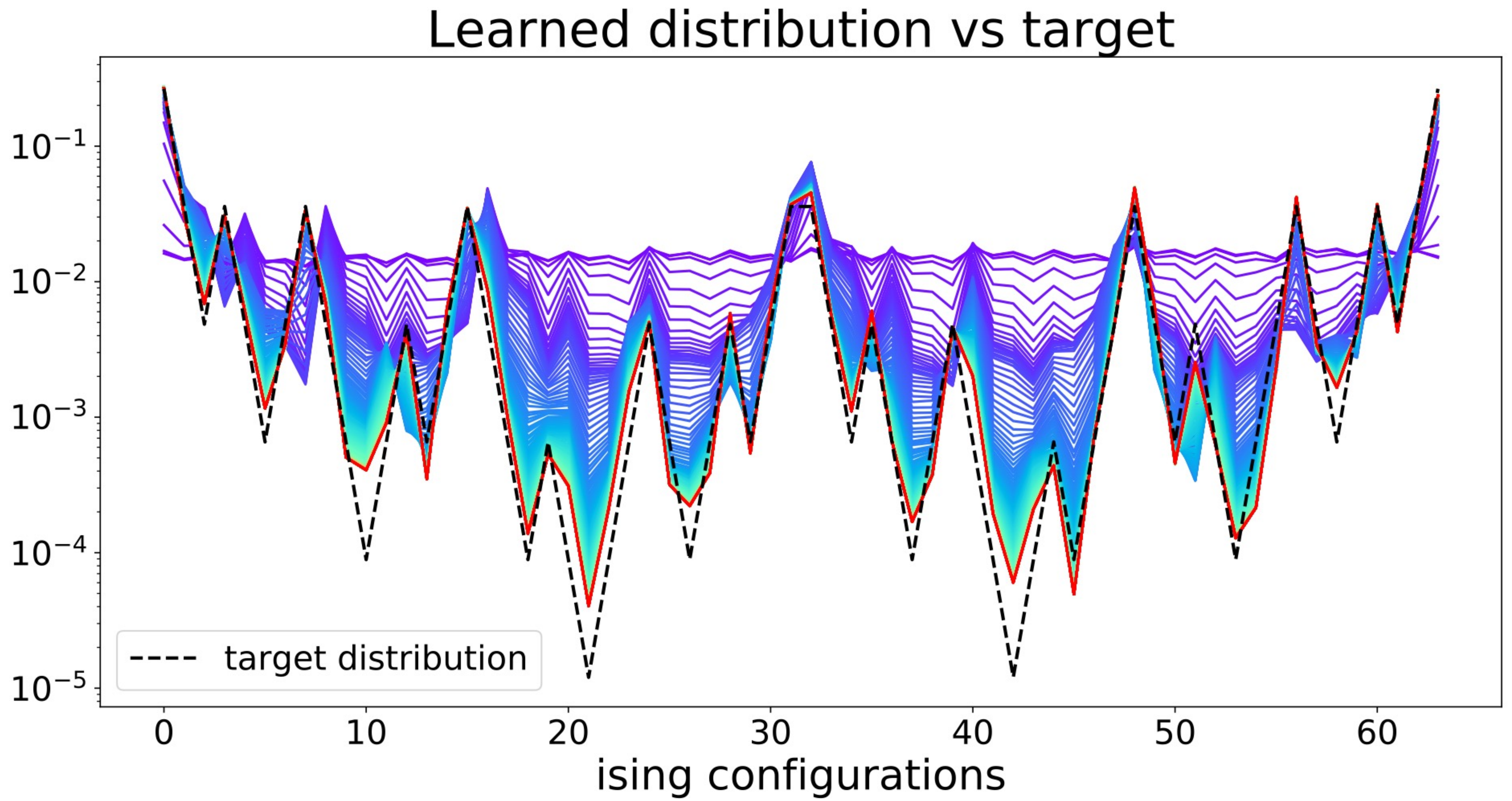
$\implies$ Contrastive Divergence

Estimate $p(h_i = 1|\mathbf{v})v_j - \langle p(h_i = 1|\mathbf{v}')v_j \rangle_{p(\mathbf{v}')}$, $\forall i, j$ as:

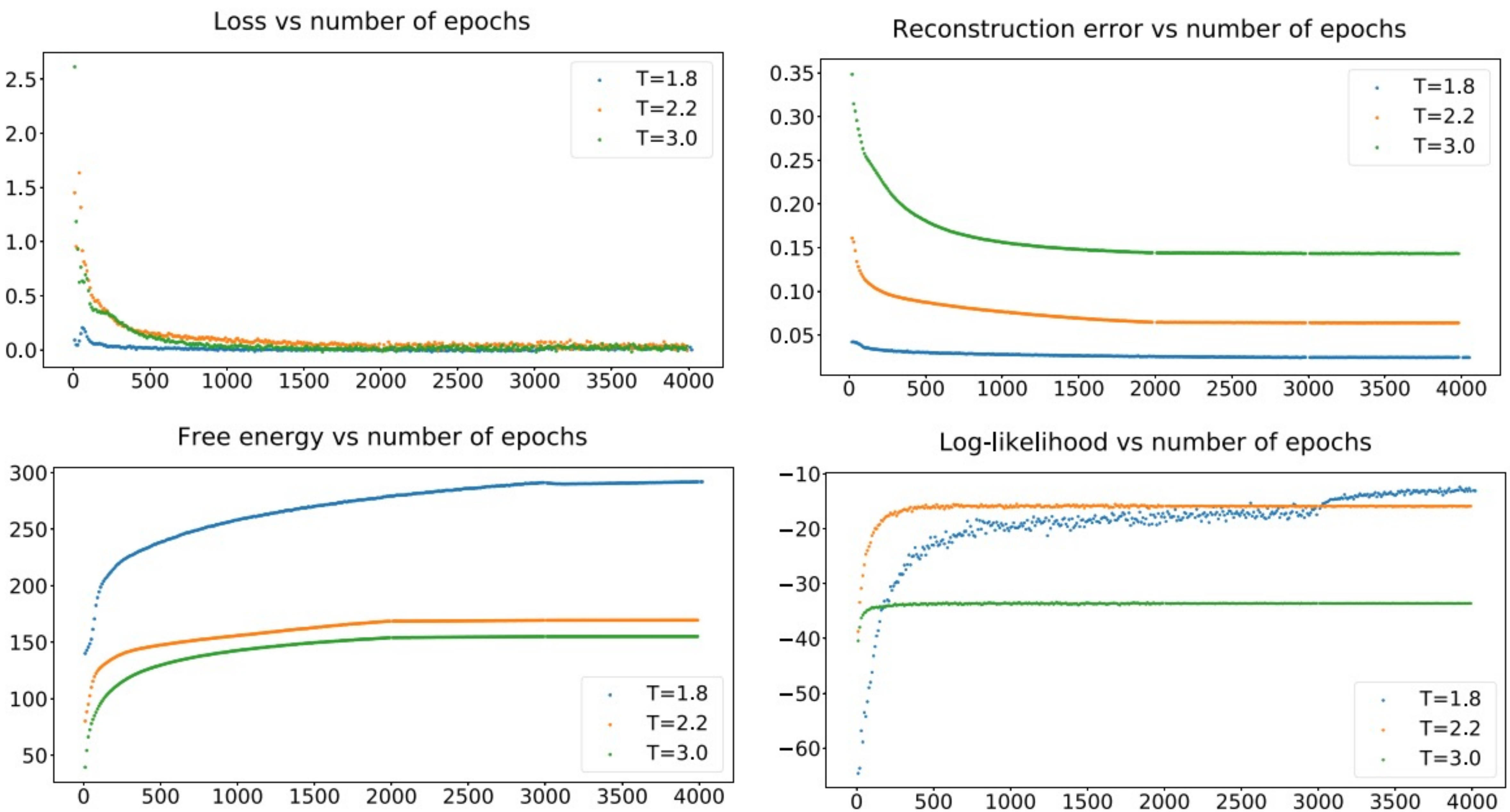
$$p(h_i = 1|\mathbf{v^{(0)}})v_j - p(h_i = 1|\mathbf{v^{(k)}})v_j^{(k)}$$

- k : Gibbs sampling step, typically set $k = 1$
- Initialised with a training example $\mathbf{v^{(0)}}$
- Each step t involves sampling $h^{(t)} \sim p_{\text{RBM}}(h_i = 1|\mathbf{v}^{(t)}, \theta)$, then sampling $v^{(t+1)} \sim p_{\text{RBM}}(v_j = 1|\mathbf{h}^{(t)}, \theta)$

RBM: 1D Ising model with 6 variables



RBM: 2D Ising, monitoring the training procedure



Cossu et. al., Physical Review B (2018)

2D Ising model observables

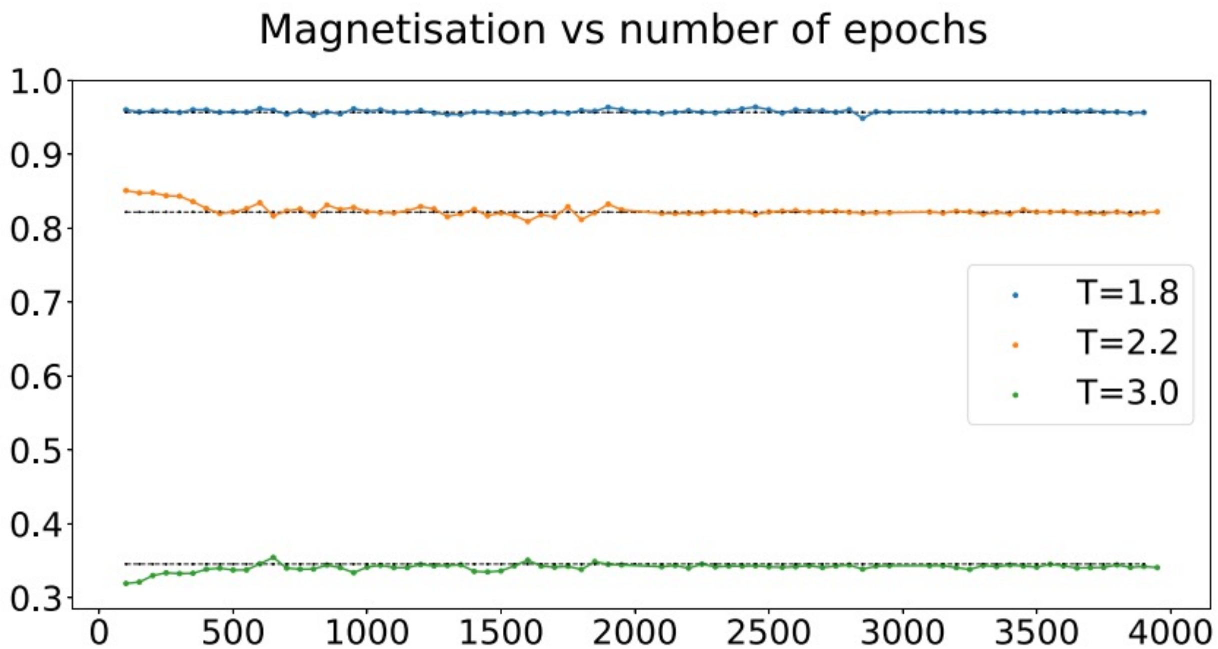
$$\langle m \rangle = \frac{1}{L^2} \left\langle \left| \sum_{i=1}^{L^2} s_i \right| \right\rangle ,$$

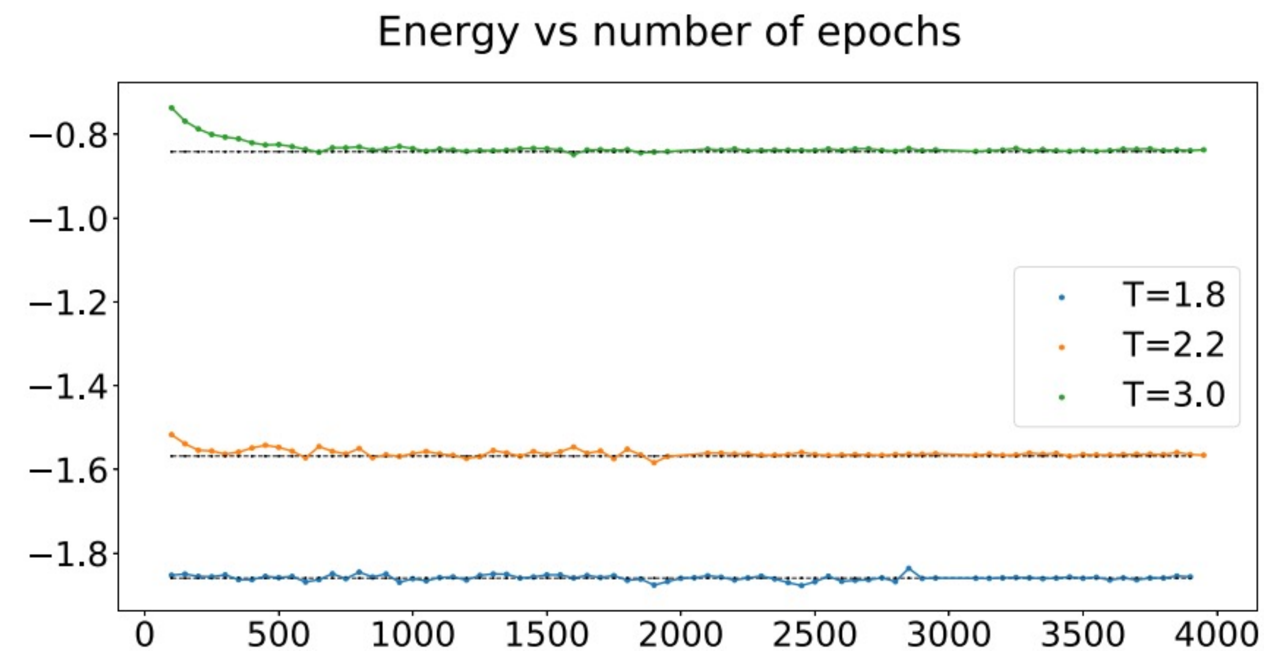
$$\langle \chi \rangle = \frac{L^2}{T} \left\langle \langle m^2 \rangle - \langle m \rangle^2 \right\rangle ,$$

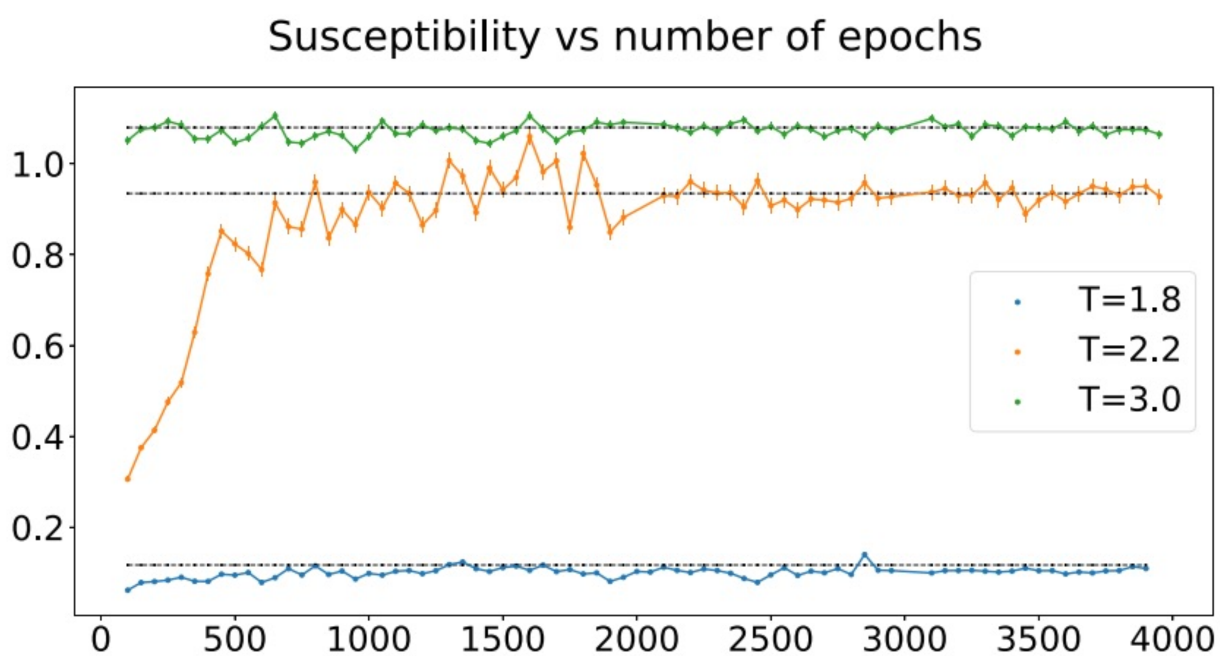
$$\langle E \rangle = -\frac{1}{L^2} \left\langle \sum_{\langle i,j \rangle} s_i s_j \right\rangle ,$$

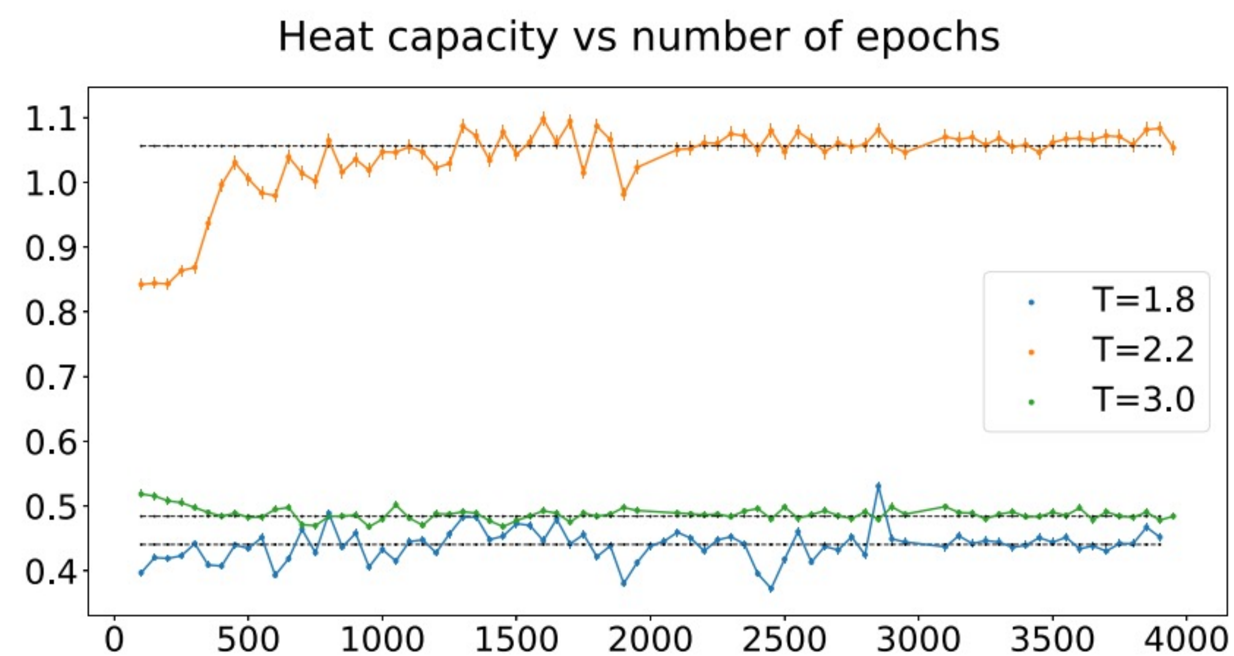
$$\langle c_v \rangle = \frac{L^2}{T^2} \left\langle \langle E^2 \rangle - \langle E \rangle^2 \right\rangle .$$

RBM: 2D Ising, monitoring the training procedure



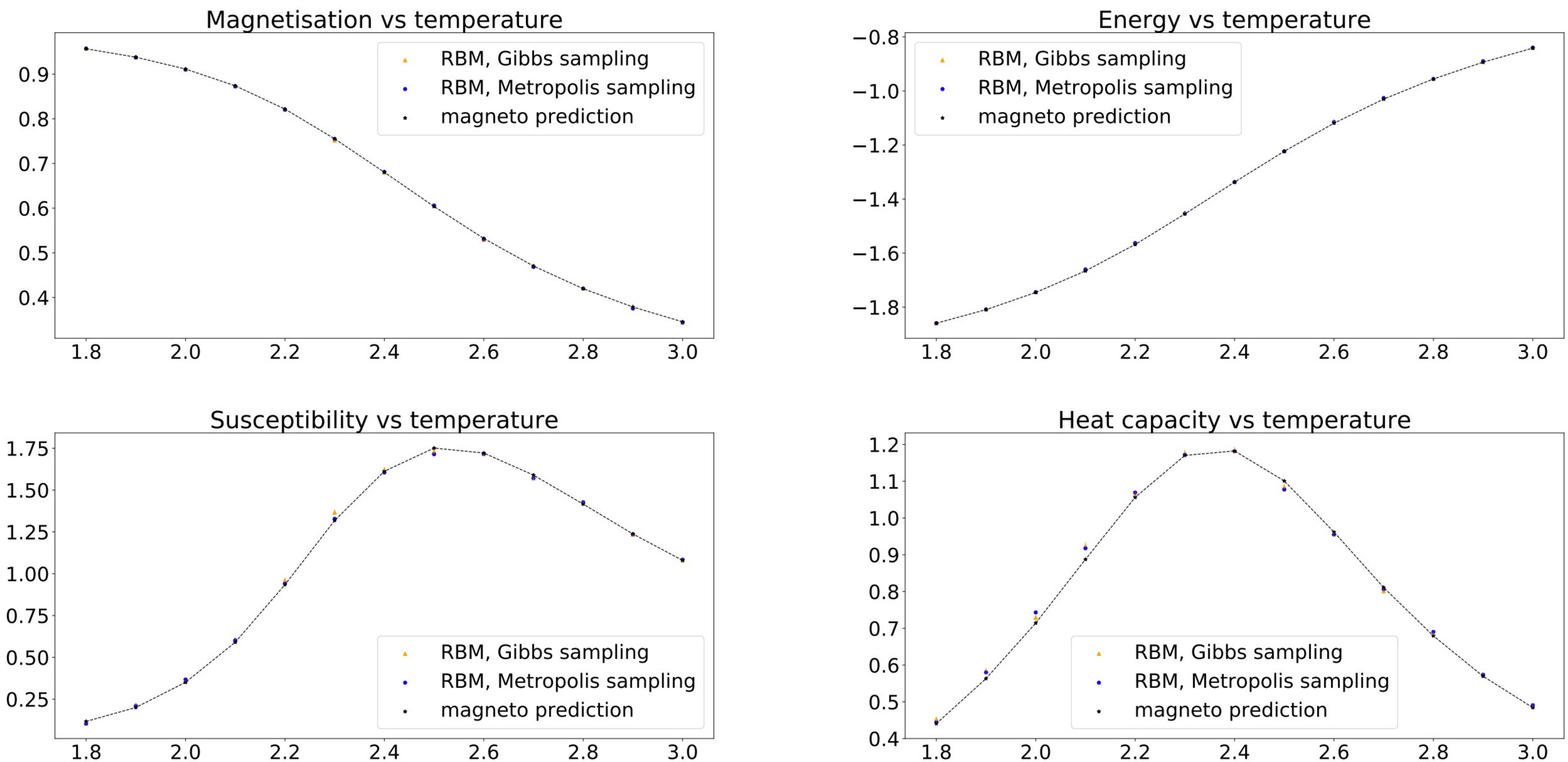






Cossu et. al., Physical Review B (2018)

RBM: 2D Ising, monitoring the training procedure



Cossu et. al., Physical Review B (2018)

RBM Prediction: n-point interactions

- A non pair-wise treatment
- Higher order couplings
- Not accessible via standard statistical techniques

$$E(\mathbf{v}) = -\sum_j b_j v_j - \sum_j \left(\sum_i \kappa_i^{(i)} W_{ij} \right) v_j - \frac{1}{2} \sum_{jk} \underbrace{\left(\sum_i \kappa_i^{(2)} W_{ik} W_{ij} \right)} v_j v_k + \cdots$$

Re-sum the <u>entire</u> series to obtain 2-point coupling!!

Derivation of n-point interactions in closed form

$$E(\mathbf{v}) = \ln \sum_{\mathbf{h}} e^{E(\mathbf{v},\mathbf{h})}$$

$$= \quad \ln \sum_{\mathbf{h}} e^{-\sum_j b_j v_j - \sum_i c_i h_i - \sum_{i,j} h_i W_{ij} v_j}$$

$$E(\mathbf{v}) = -\sum_j b_j v_j - \sum_i \ln \sum_{h_i} e^{c_i h_i} e^{\sum_j h_i W_{ij} v_j}$$

$$= -\sum_j b_j v_j - \sum_i \ln \sum_{h_i} q(h_i) e^{t h_i} \ ,$$

$t \equiv \sum_j W_{ij} v_j$ and $q(h_i) \equiv e^{c_i h_i}$

Cumulant generating function:

$$K_i(t) \equiv \ln \sum_{h_i} q(h_i) e^{t h_i} = \sum_n \frac{\kappa_i^{(n)} t^n}{n!}$$

$$\kappa_i^{(n)} = \partial_t^n K_i(t)|_{t=0}$$

A high-bias, low-variance introduction to machine learning for physicists by Mehta et al. *Physics Reports* (2019)

Derivation of n-point interactions in closed form

$$
\begin{aligned}
E(\mathbf{v}) &= -\sum_j b_j v_j - \sum_i \kappa_i^{(0)} - \sum_i \kappa_i^{(1)} t - \sum_i \frac{\kappa_i^{(2)} t^2}{2!} - \dots \\
&= -\sum_i \kappa_i^{(0)} - \sum_j \left(b_j + \sum_i \kappa_i^{(1)} W_{ij} \right) v_j - \frac{1}{2!} \sum_{j_1, j_2} \left(\sum_i \kappa_i^{(2)} W_{ij_1} W_{ij_2} \right) v_{j_1} v_{j_2} - \dots
\end{aligned}
$$

$$
\boxed{v_j^n = v_j \quad , \quad n \in \mathbb{Z}^+}
$$

e.g. 2-point interaction:

$$
\sum_{n>1} \frac{1}{2(n!)} \sum_{0<k<n} \sum_{j_1 \neq j_2} \left(\sum_i \kappa_i^{(n)} \binom{n}{k} W_{ij_1}^{k} W_{ij_2}^{n-k} \right) v_{j_1} v_{j_2}
$$

...

$$
H_{j_1 j_2} = \frac{1}{8} \sum_i \ln \frac{(1 + e^{c_i + W_{ij_1} + W_{ij_2}})(1 + e^{c_i})}{(1 + e^{c_i + W_{ij_1}})(1 + e^{c_i + W_{ij_2}})}
$$

Closed form expression!

Cossu et. al., Physical Review B (2018)

Derivation of n-point interactions in closed form

$$E(\mathbf{v}) = -\sum_j b_j v_j - \sum_i \kappa_i^{(0)} - \sum_i \kappa_i^{(1)} t - \sum_i \frac{\kappa_i^{(2)} t^2}{2!} - \dots$$
$$= -\sum_i \kappa_i^{(0)} - \sum_j \left(b_j + \sum_i \kappa_i^{(1)} W_{ij} \right) v_j - \frac{1}{2!} \sum_{j_1, j_2} \left(\sum_i \kappa_i^{(2)} W_{ij_1} W_{ij_2} \right) v_{j_1} v_{j_2} - \dots$$

$$\boxed{v_j^n = v_j \quad , \quad n \in \mathbb{Z}^+}$$

e.g. 2-point interaction:

$$\sum_{n>1} \frac{1}{2(n!)} \sum_{0<k<n} \sum_{j_1 \neq j_2} \left(\sum_i \kappa_i^{(n)} \binom{n}{k} W_{ij_1}^{k} W_{ij_2}^{n-k} \right) v_{j_1} v_{j_2}$$

e.g. 3-point interaction: ...

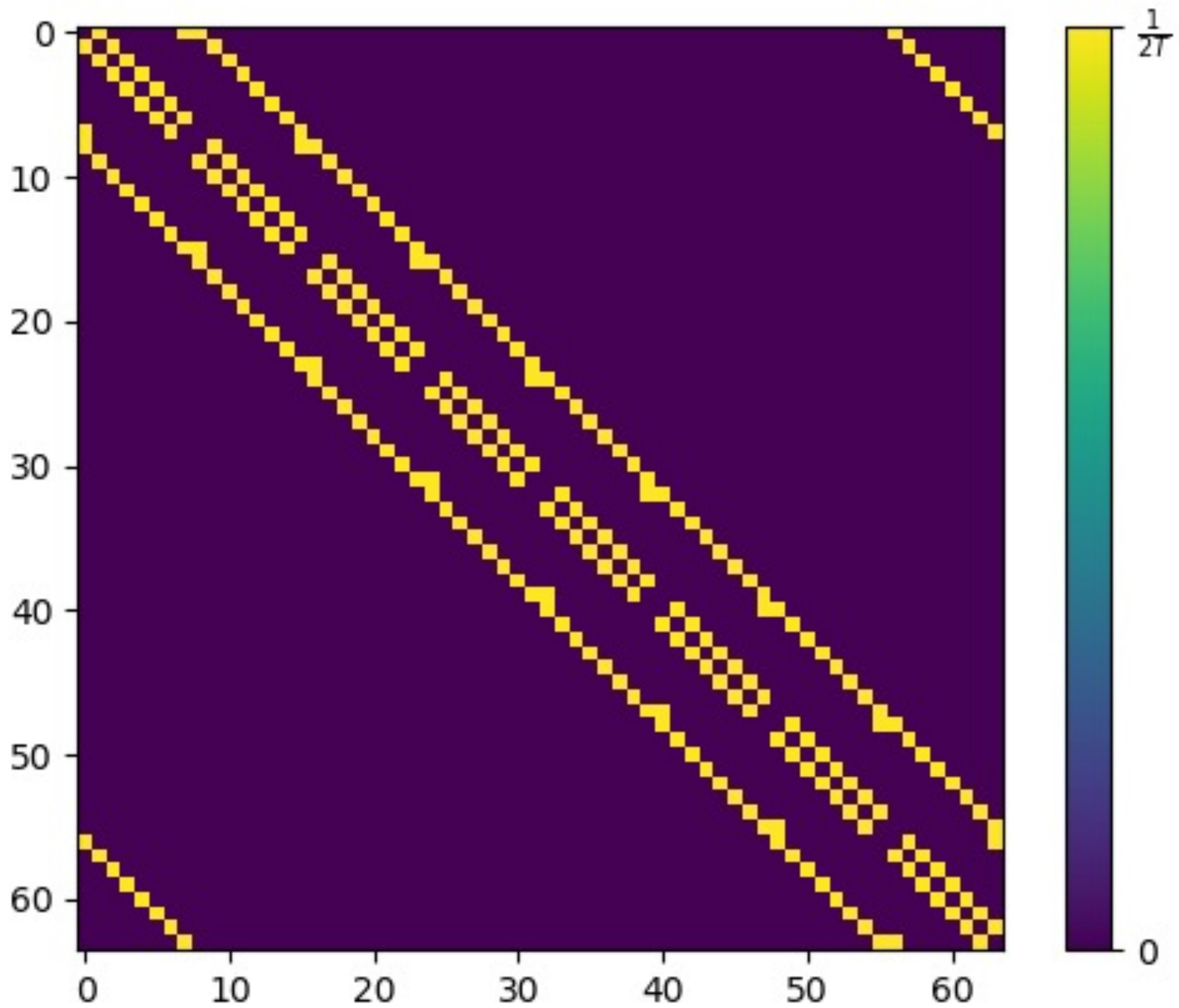
$$\frac{1}{6} \sum_i \ln \frac{(1+e^{c_i + W_{ij_1} + W_{ij_2} + W_{ij_3}})(1+e^{c_i + W_{ij_1}})(1+e^{c_i + W_{ij_2}})(1+e^{c_i + W_{ij_3}})}{(1+e^{c_i + W_{ij_1} + W_{ij_2} +})(1+e^{c_i + W_{ij_1} + W_{ij_3}})(1+e^{c_i + W_{ij_2} + W_{ij_3}})(1+e^{c_i})}$$

Closed form expression!

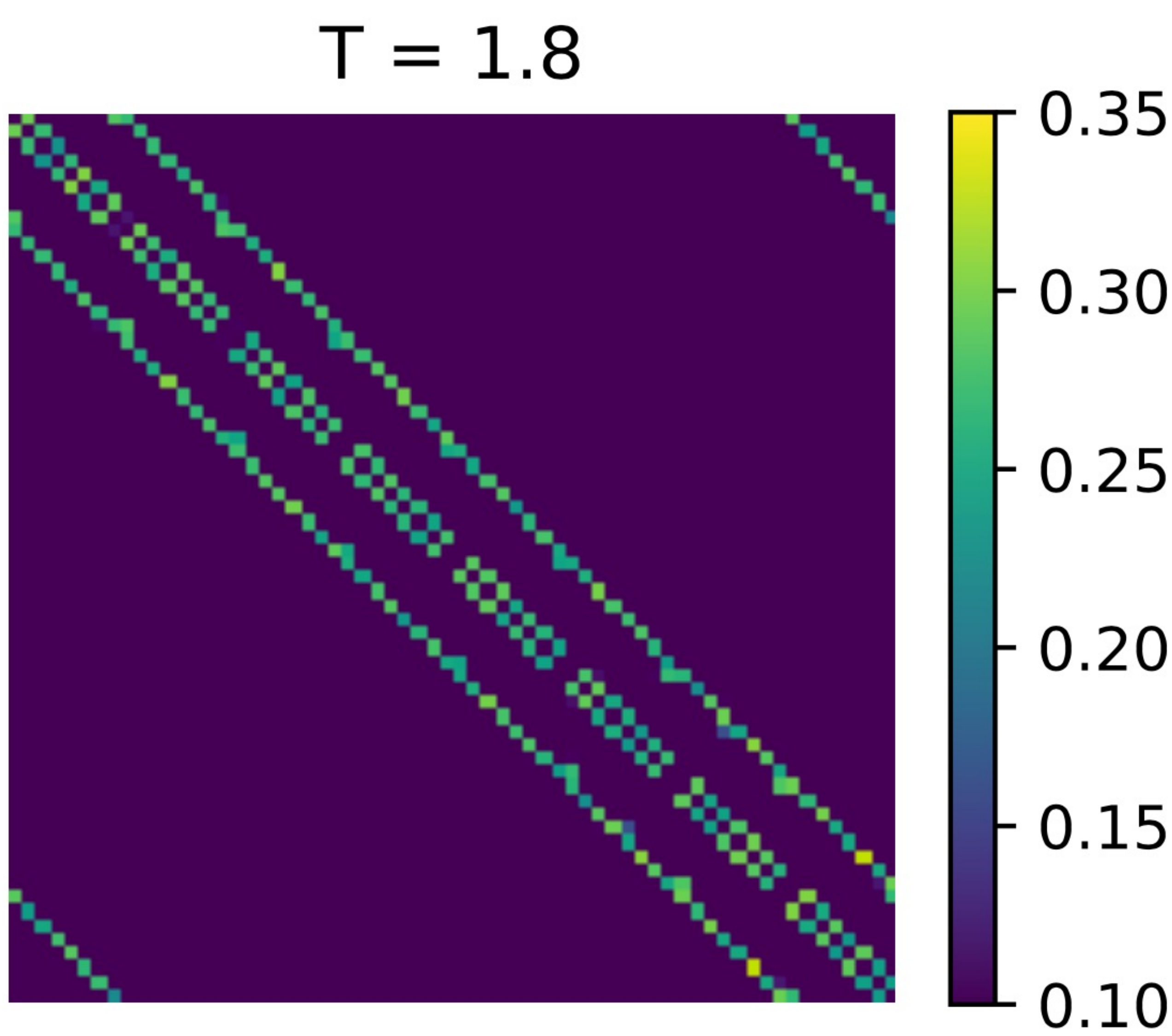
Cossu et. al., Physical Review B (2018)

RBM Predictions: Couplings J_{ij}

Ising Model

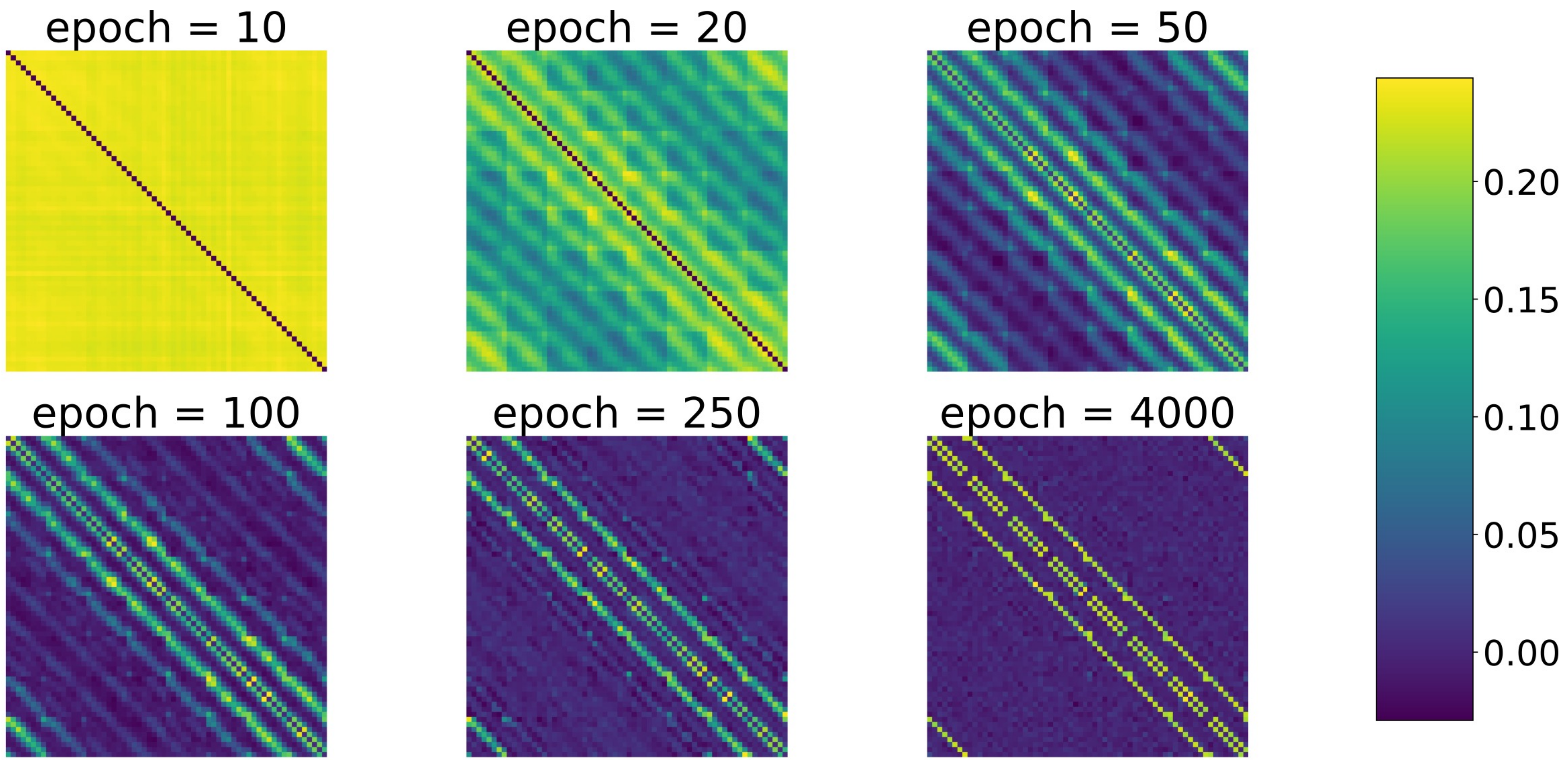


RBM prediction

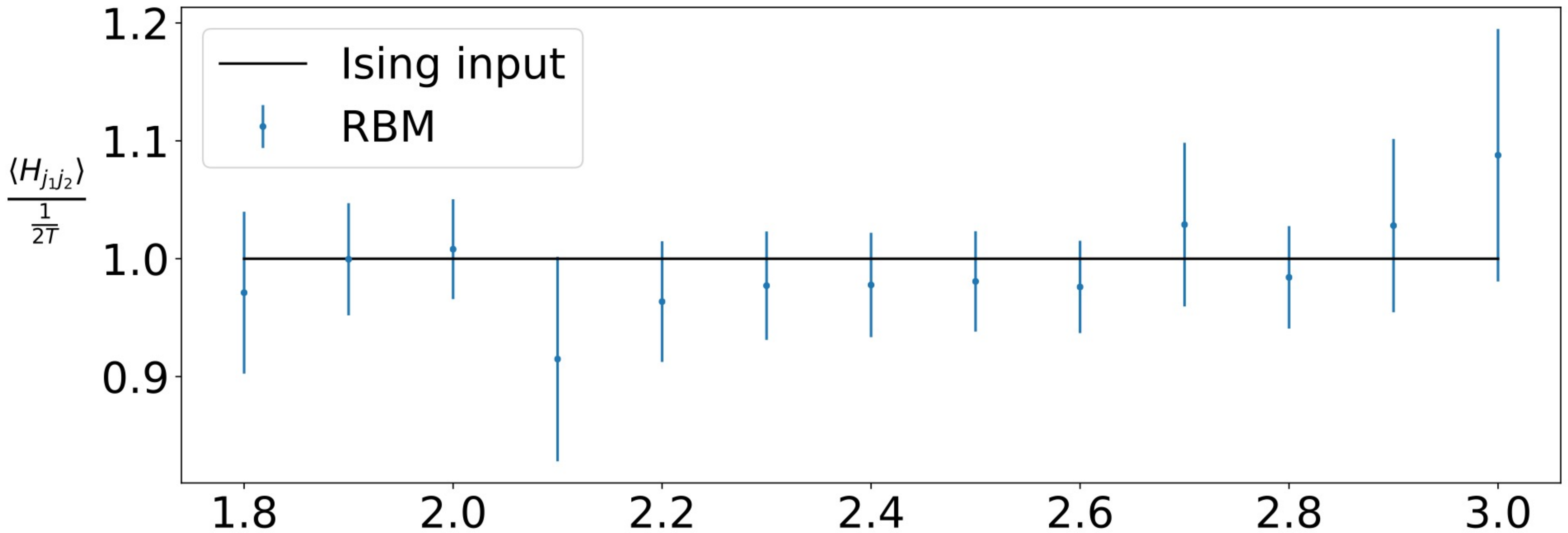


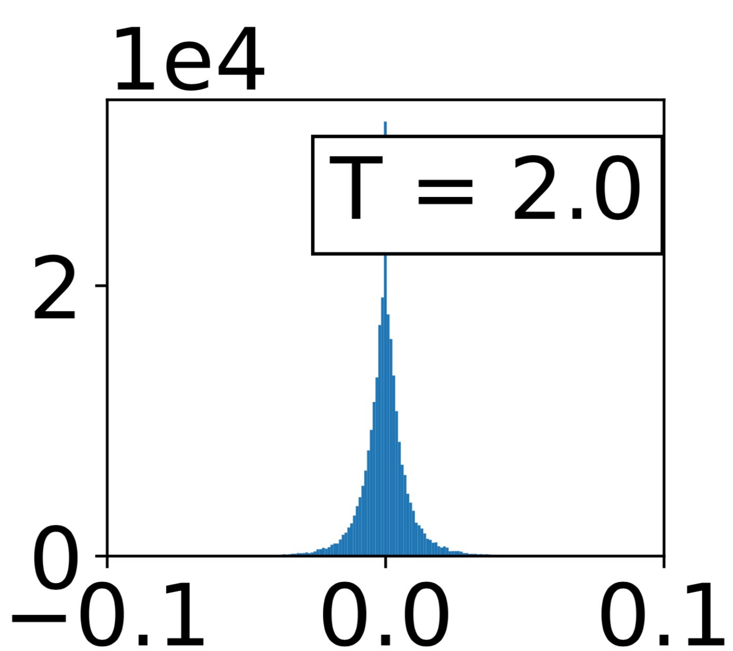
Cossu et. al., Physical Review B (2018)

Couplings during training

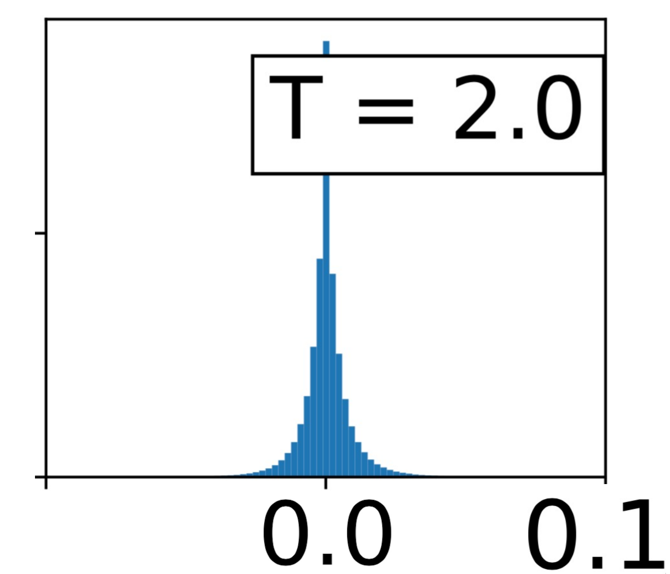


RBM Predictions: Couplings (normalised)

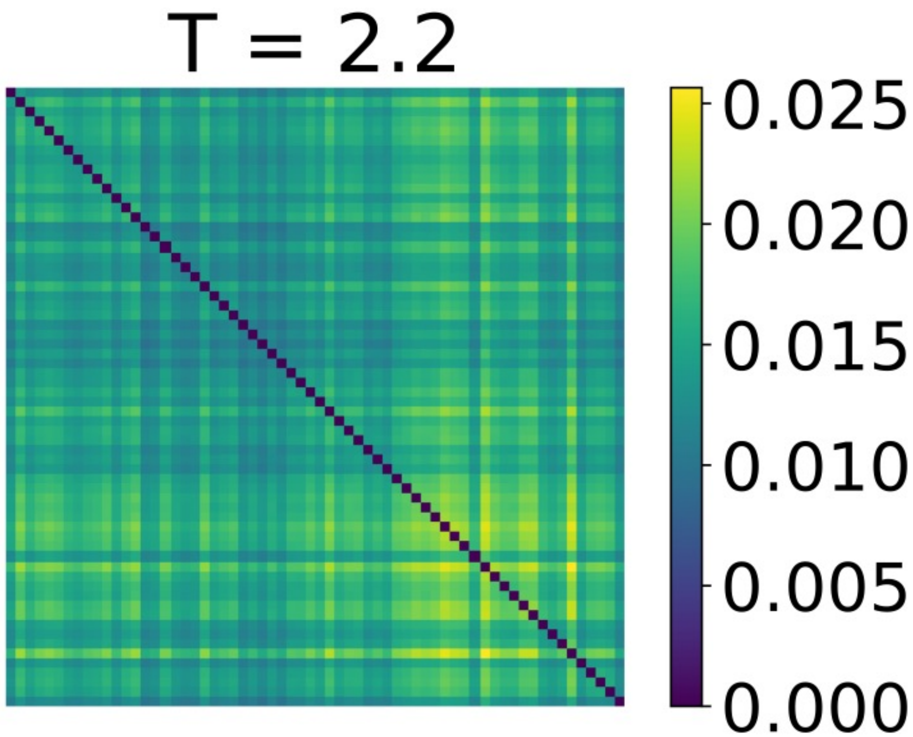




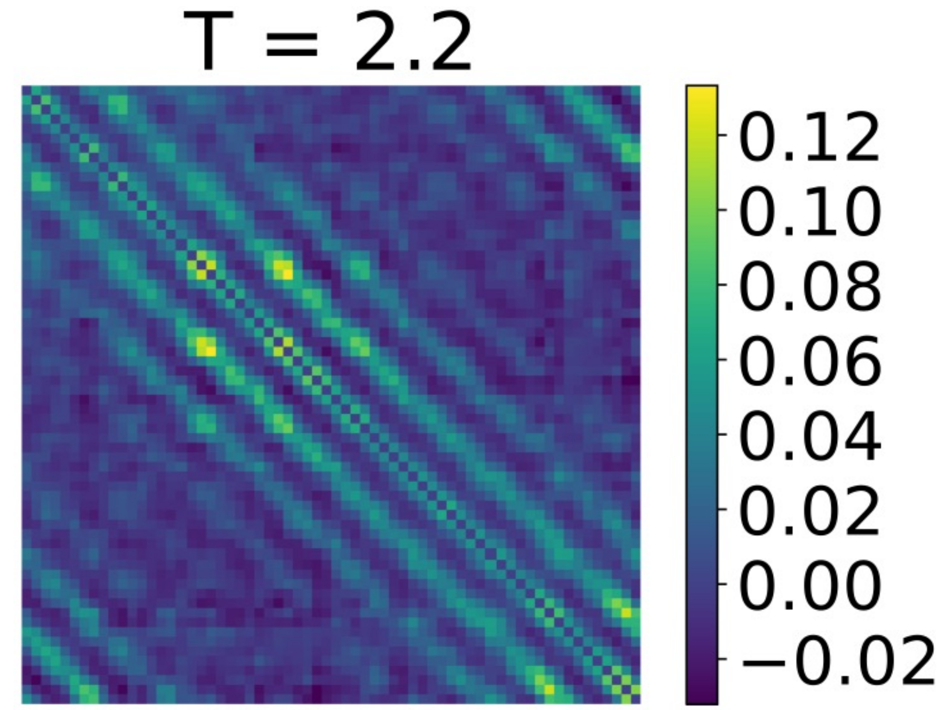
3- and 4-point couplings



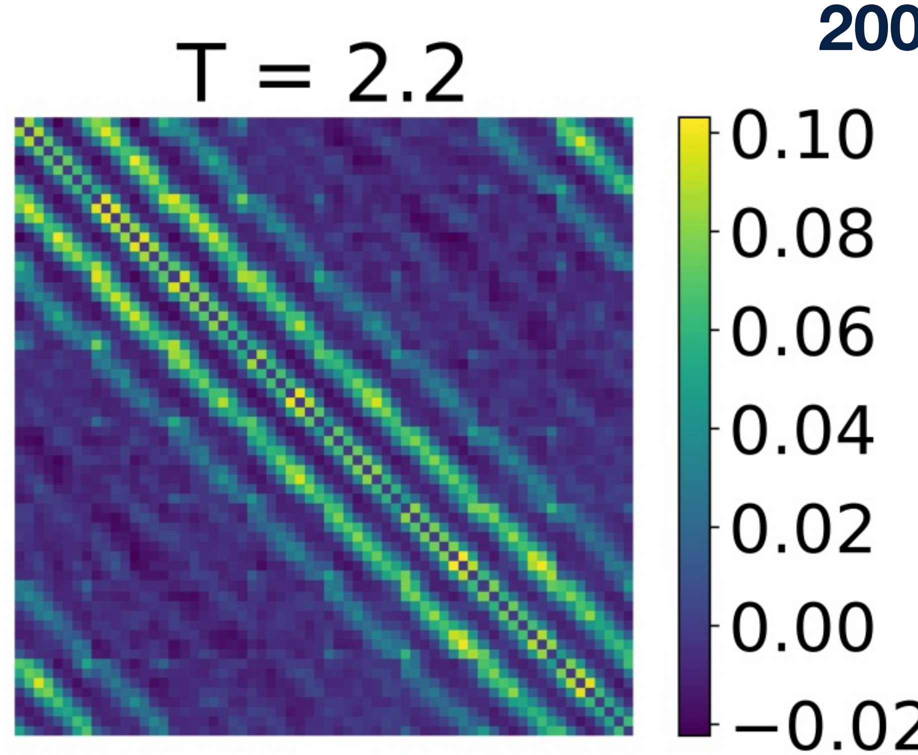
Small number of training examples



200 Examples



2000 Examples



10000 Examples

Lessons Learnt

- Understand well the training criteria from RBMs: Log-likelihood, Loss, free energy, reconstruction error + moments generated by the machine

Lessons Learnt

- Understand well the training criteria from RBMs: Log-likelihood, Loss, free energy, reconstruction error + moments generated by the machine

- RBMs are successful at prediction (higher-order) interactions in a given system of binary variables

Lessons Learnt

- Understand well the training criteria from RBMs: Log-likelihood, Loss, free energy, reconstruction error + moments generated by the machine

- RBMs are successful at prediction (higher-order) interactions in a given system of binary variables

- Generally, need lots of training examples

Lessons Learnt

- Understand well the training criteria from RBMs: Log-likelihood, Loss, free energy, reconstruction error +
  moments generated by the machine

- RBMs are successful at prediction (higher-order) interactions in a given system of binary variables

- Generally, need lots of training examples

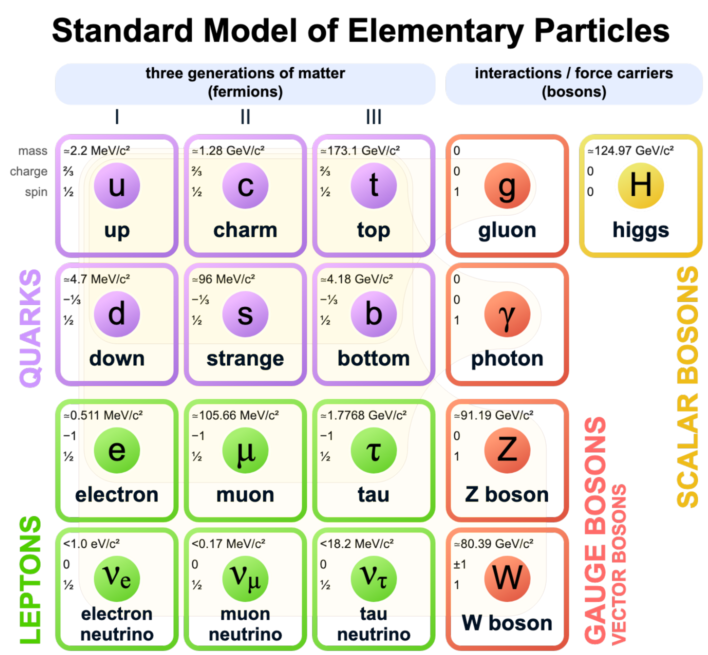
AND

- Still need to deal with potentially very large numbers of variables (e.g. Gene Networks)
- RBMs are not particularly convenient to train … (e.g. including time on hyper parameter tuning)

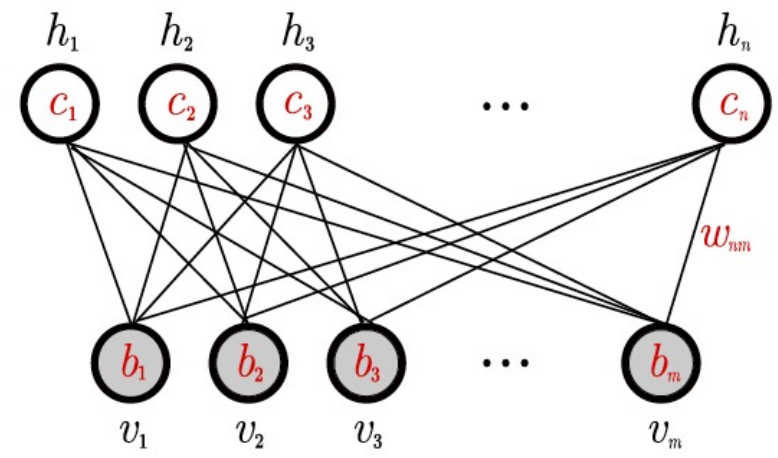
Interactions:
Model-independent definition
&
Estimation

Interaction in science

Elementary Particles

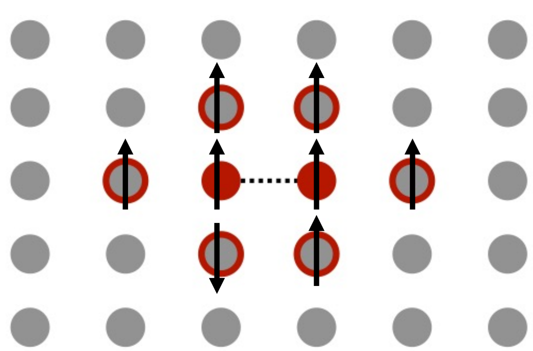


Nodes in a neural network

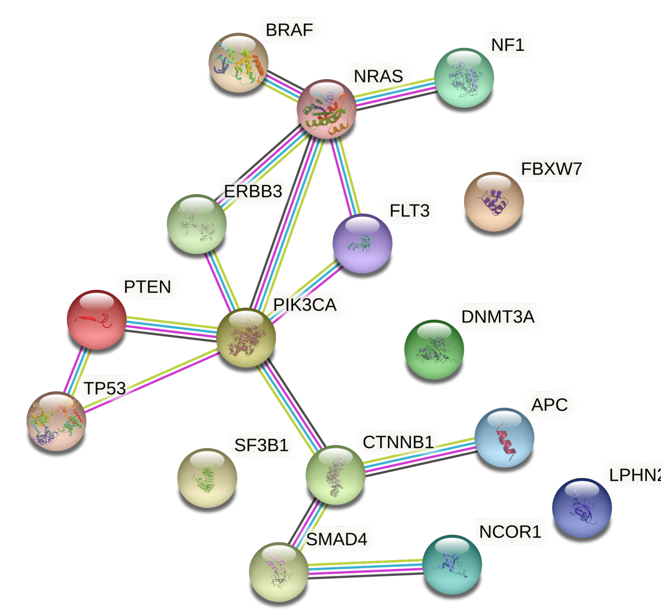


Interactions

Statistical Physics



Biomedicine



From the Question to defining the target quantity of interest

Aim: Formulate the **target** quantity of interest:

not as a property of a parametric statistical model

The target quantity can often be identified **without** ever specifying the functional or distributional form of the model: **model-independent**

Judea Pearl (2010), Mark van der Laan (2011)

From the Question to defining the target quantity of interest

Aim: Formulate the **target** quantity of interest:

not as a property of a parametric statistical model

The target quantity can often be identified **without** ever specifying the functional or distributional form of the model: **model-independent**

Targeted Learning

Why is this important?

1) Be clear about what we are actually after.
2) Don't waste computational, analytical and data resources on irrelevant aspects of a problem
3) Focus on what is relevant: answering questions we actually care about!

Judea Pearl (2010), Mark van der Laan (2011)

Example

Define: Express the target quantity of interest, **interaction**, as a function that can be computed for ***any*** **model**, i.e. model-independent

$$\mathbb{E}_0\left(G_3 | G_1, G_2\right) = \alpha_0 + \alpha_1 G_1 + \alpha_2 G_2 + \gamma G_1 G_2$$

Example

Define: Express the target quantity of interest, **interaction**, as a function that can be computed for ***any*** **model**, i.e. model-independent

$$\mathbb{E}_0\left(G_3|G_1,G_2\right) = \alpha_0 + \alpha_1 G_1 + \alpha_2 G_2 + \gamma G_1 G_2$$

$$\mathbb{E}_0\left(G_3|G_1,G_2\right) = \alpha_0 + \alpha_1 G_1 + \alpha_2 G_2 + \underline{\gamma} G_1 G_2 + \underline{\gamma'} G_1^2 G_2$$

Example

Define: Express the target quantity of interest, **interaction**, as a function that can be computed for ***any*** **model**, i.e. model-independent

$$\mathbb{E}_0\left(G_3|G_1,G_2\right) = \alpha_0 + \alpha_1 G_1 + \alpha_2 G_2 + \gamma G_1 G_2$$

$$\mathbb{E}_0\left(G_3|G_1,G_2\right) = \alpha_0 + \alpha_1 G_1 + \alpha_2 G_2 + \gamma G_1 G_2 + \gamma' G_1^2 G_2$$

$$\mathbb{E}_0\left(G_3|G_1,G_2\right) = \frac{1}{1+e^{\alpha_0+\alpha_1 G_1+\alpha_2 G_2+\underline{\gamma} G_1 G_2}}$$

Example

Define: Express the target quantity of interest, **interaction**, as a function that can be computed for ***any*** **model**, i.e. model-independent

$$\mathbb{E}_0\left(G_3|G_1,G_2\right) = \alpha_0 + \alpha_1 G_1 + \alpha_2 G_2 + \gamma G_1 G_2$$

$$\mathbb{E}_0\left(G_3|G_1,G_2\right) = \alpha_0 + \alpha_1 G_1 + \alpha_2 G_2 + \gamma G_1 G_2 + \gamma' G_1^2 G_2$$

$$\mathbb{E}_0\left(G_3|G_1,G_2\right) = \frac{1}{1+e^{\alpha_0+\alpha_1 G_1+\alpha_2 G_2+\gamma G_1 G_2}}$$

$$\mathbb{E}_0\left(G_3|G_1,G_2\right) = \frac{1}{1+e^{\alpha_0+\alpha_1 G_1+\alpha_2 G_2}}$$

???

Model-independent formulation of interactions

Key: function that can be computed for ***any* model**. Function of the distribution without needing to specify its parametric form.

Average Treatment Effect (ATE):

$$\mathrm{ATE}_G(Y) = \mathbb{E}_W\left[\mathbb{E}(Y \mid G = 1, W) - \mathbb{E}(Y \mid G = 0, W)\right]$$

Outcome:
-Cancer/not
-Health outcome

Treatment:
-mutation/not
-Drug 0 or 1

Confounder

Model-independent formulation of interactions

Key: function that can be computed for ***any*** **model**. Function of the distribution without needing to specify its parametric form.

Average Treatment Effect (ATE):

$$\mathrm{ATE}_G(Y) = \mathbb{E}_W\left[\mathbb{E}(Y \mid G=1, W) - \mathbb{E}(Y \mid G=0, W)\right]$$

Outcome:
-Cancer/not
-Health outcome

Treatment:
-mutation/not
-Drug 0 or 1

Confounder

Interactions between variables i and j leading to outcome Y:

$$\begin{aligned} I^a_{i,j} = & \left[\mathbb{E}(Y \mid G_{ij} = (1,1), \underline{G} = 0) - \mathbb{E}(Y \mid G_{ij} = (0,1), \underline{G} = 0)\right] \\ & - \left[\mathbb{E}(Y \mid G_{ij} = (1,0), \underline{G} = 0) - \mathbb{E}(Y \mid G_{ij} = (0,0), \underline{G} = 0)\right]. \end{aligned}$$

Beentjes & Khamseh, Physical Review E (2020)

Model-independent formulation of interactions

Key: function that can be computed for ***any*** **model**. Function of the distribution without needing to specify its parametric form.

Average Treatment Effect (ATE):

$$\mathrm{ATE}_G(Y) = \mathbb{E}(Y \mid G = 1) - \mathbb{E}(Y \mid G = 0)$$

Interactions between variables i and j leading to outcome Y:

$$\begin{aligned} I^a_{i,j} = & \boxed{\big[\mathbb{E}(Y \mid G_{ij} = (1,1), \underline{G} = 0) - \mathbb{E}(Y \mid G_{ij} = (0,1), \underline{G} = 0)\big]} \\ & - \big[\mathbb{E}(Y \mid G_{ij} = (1,0), \underline{G} = 0) - \mathbb{E}(Y \mid G_{ij} = (0,0), \underline{G} = 0)\big]. \end{aligned}$$

Spin i is up (1) or down (0)

Spin j up (1)

Beentjes & Khamseh, Physical Review E (2020)

Model-independent formulation of interactions

Key: function that can be computed for ***any*** **model**. Function of the distribution without needing to specify its parametric form.

Average Treatment Effect (ATE):

$$\mathrm{ATE}_G(Y) = \mathbb{E}(Y \mid G = 1) - \mathbb{E}(Y \mid G = 0)$$

Interactions between variables i and j leading to outcome Y:

$$\begin{aligned} I^a_{i,j} = &\left[\mathbb{E}(Y \mid G_{ij} = (1,1), \underline{G} = 0) - \mathbb{E}(Y \mid G_{ij} = (0,1), \underline{G} = 0)\right] \\ &- \left[\mathbb{E}(Y \mid G_{ij} = (1,0), \underline{G} = 0) - \mathbb{E}(Y \mid G_{ij} = (0,0), \underline{G} = 0)\right]. \end{aligned}$$

Spin i is up (1) or down (0)

Spin j <u>down</u> (0)

Model-independent formulation of interactions

Key: function that can be computed for ***any*** **model**. Function of the distribution without needing to specify its parametric form.

Average Treatment Effect (ATE):

$$\mathrm{ATE}_G(Y) = \mathbb{E}(Y \mid G = 1) - \mathbb{E}(Y \mid G = 0)$$

Interactions between variables i and j leading to outcome Y:

$$\begin{aligned} I^a_{i,j} = & \big[\mathbb{E}(Y \mid G_{ij} = (1,1), \underline{G} = 0) - \mathbb{E}(Y \mid G_{ij} = (0,1), \underline{G} = 0)\big] \\ & - \big[\mathbb{E}(Y \mid G_{ij} = (1,0), \underline{G} = 0) - \mathbb{E}(Y \mid G_{ij} = (0,0), \underline{G} = 0)\big]. \end{aligned}$$

"Does variable i influence outcome differently, depending on the status of variable j?"

$$\boxed{I^a_{i,j} = \partial_{G_i} \partial_{G_j} \mathbb{E}\left(Y \middle| G_1, \ldots, G_N\right)}$$

Beentjes & Khamseh, Physical Review E (2020)

Model-independent formulation of interactions

Key: function that can be computed for ***any*** **model**. Function of the distribution without needing to specify its parametric form.

Average Treatment Effect (ATE):

$$\mathrm{ATE}_G(Y) = \mathbb{E}(Y \mid G = 1) - \mathbb{E}(Y \mid G = 0)$$

Interactions between variables i and j leading to outcome Y:

$$\begin{aligned} I^a_{i,j} = & \big[\mathbb{E}(Y \mid G_{ij} = (1,1), \underline{G} = 0) - \mathbb{E}(Y \mid G_{ij} = (0,1), \underline{G} = 0)\big] \\ & - \big[\mathbb{E}(Y \mid G_{ij} = (1,0), \underline{G} = 0) - \mathbb{E}(Y \mid G_{ij} = (0,0), \underline{G} = 0)\big]. \end{aligned}$$

Generalised to higher-order interaction, e.g.,

"Does the interaction between variable i and variable j influence outcome differently, depending on the status of variable k?"

Beentjes & Khamseh, Physical Review E (2020)

Model-independent formulation of interactions

Key: function that can be computed for ***any*** **model**. Function of the distribution without needing to specify its parametric form.

Average Treatment Effect (ATE):

$$\mathrm{ATE}_G(Y) = \mathbb{E}(Y \mid G = 1) - \mathbb{E}(Y \mid G = 0)$$

All order interactions:

$$\boxed{I^a_{i_1,\ldots,i_n} = \sum_{k=0}^{n} (-1)^{n-k} \left(\sum_{J \subset I \,:\, \ell(J) = k} \mathbb{E}\big(Y \mid G_I = e_J^{(n)}, \underline{G} = 0\big) \right)}$$

Where $e_J^{(\ell(I))} = (e_{i_1}, \ldots, e_{i_{\ell(I)}})$ is a tuple of elements taking values 0 or 1 (k)

Beentjes & Khamseh, Physical Review E (2020)

Example: Linear regression

Recall: Our definition of interaction is non-parametric and model-independent

However, when applied to a particular parametric fit p_θ , we obtain the expression of n-point interaction in that parametric model (in terms of θ)

$$Y = \alpha_0 + \alpha_1 T_1 + \alpha_2 T_2 + \gamma T_1 T_2$$

$$\mathbb{E}(Y \mid T_1 = 1, T_2 = 1) = \alpha_0 + \alpha_1 + \alpha_2 + \gamma$$
$$\mathbb{E}(Y \mid T_1 = 1, T_2 = 0) = \alpha_0 + \alpha_1$$
$$\mathbb{E}(Y \mid T_1 = 0, T_2 = 1) = \alpha_0 + \alpha_2$$
$$\mathbb{E}(Y \mid T_1 = 0, T_2 = 0) = \alpha_0$$

$$\mathrm{ATE}_{T_1}(Y \mid T_2 = 1) = \alpha_1 + \gamma, \qquad \mathrm{ATE}_{T_2}(Y \mid T_1 = 1) = \alpha_2 + \gamma,$$
$$\mathrm{ATE}_{T_1}(Y \mid T_2 = 0) = \alpha_1 \qquad \mathrm{ATE}_{T_2}(Y \mid T_1 = 0) = \alpha_2$$

$$\boxed{I^a_{1,2} = \gamma = I^a_{2,1}}$$

Numerical example: Linear regression

$$Y = \alpha_0 + \alpha_1 T_1 + \alpha_2 T_2 + \alpha_3 T_3 + \alpha_{12} T_1 T_2 + \alpha_{13} T_1 T_3 + \alpha_{23} T_2 T_3 + \gamma T_1 T_2 T_3 + \epsilon$$

Without loss of generality, set $\gamma = 2\sigma^2 = 2$ $\quad \alpha_{12}, \alpha_{13}, \alpha_{23} = 5.0, -2.5, 0$

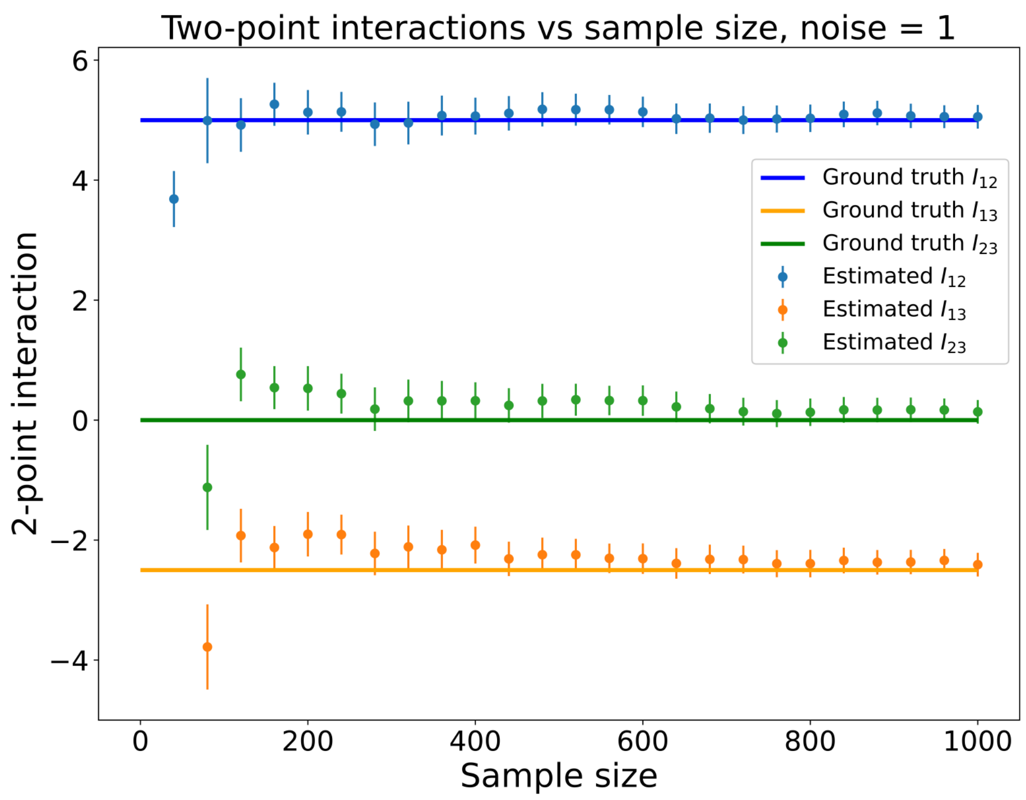
$\alpha_1, \alpha_2, \alpha_3 = -2, 10, 0$, $\alpha_0 = -1.5$

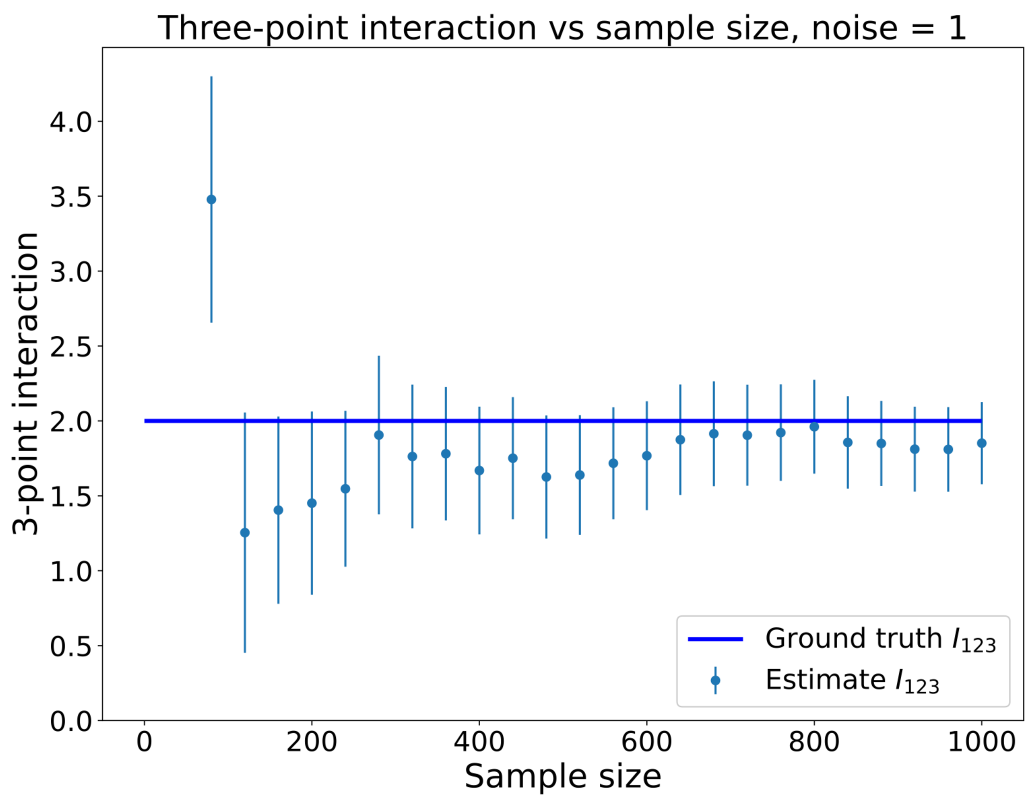
Generate data: $\epsilon \sim \mathcal{N}(0, \sigma^2)$, $\sigma^2 = 1$, $T_1 \sim \text{Binom}(p = 0.4)$

$T_2 \sim \text{Binom}(p = 0.7)$, $T_3 \sim \text{Binom}(p = 0.5)$

Obtain estimates using the TL additive formulation

Numerical example: Linear regression





Multiplicative formulation of interactions

Alternatively: Do not wish to specify an outcome, instead ask how spins being up/down influence their joint probability? e.g. interactions amongst spins in a network

Start with 1 spin: $$I_i^m = \ln\left(\frac{p(G_i = 1 \mid \underline{G} = 0)}{p(G_i = 0 \mid \underline{G} = 0)}\right)$$

'odds ratio': What is the likelihood of spin i being 1 vs 0

Beentjes & Khamseh, Physical Review E (2020)

Multiplicative formulation of interactions

Alternatively: Do not wish to specify an outcome, instead ask how spins being up/down influence their joint probability? e.g. interactions amongst spins in a network

Start with 1 spin:

$$I_i^m = \ln\left(\frac{p(G_i = 1 \mid \underline{G} = 0)}{p(G_i = 0 \mid \underline{G} = 0)}\right)$$

'odds ratio': What is the likelihood of spin i being 1 vs 0

$$I_{i,j}^m = \ln\left(\frac{p(G_{ij} = (1,1) \mid \underline{G} = 0)}{p(G_{ij} = (0,1) \mid \underline{G} = 0)}\right) - \ln\left(\frac{p(G_{ij} = (1,0) \mid \underline{G} = 0)}{p(G_{ij} = (0,0) \mid \underline{G} = 0)}\right)$$

Multiplicative formulation of interactions

Alternatively: Do not wish to specify an outcome, instead ask how spins being up/down influence their joint probability? e.g. interactions amongst spins in a network

Start with 1 spin: $$I_i^m = \ln\left(\frac{p(G_i = 1 \mid \underline{G} = 0)}{p(G_i = 0 \mid \underline{G} = 0)}\right)$$

'odds ratio': What is the likelihood of spin i being 1 vs 0

$$I_{i,j}^m = \ln\boxed{\left(\frac{p(G_{ij} = (1,1) \mid \underline{G} = 0)}{p(G_{ij} = (0,1) \mid \underline{G} = 0)}\right)} - \ln\boxed{\left(\frac{p(G_{ij} = (1,0) \mid \underline{G} = 0)}{p(G_{ij} = (0,0) \mid \underline{G} = 0)}\right)}$$

'odd ratio' of spin i with spin j being 1

'odd ratio' of spin i with spin j being 0

Multiplicative formulation of interactions

Alternatively: Do not wish to specify an outcome, instead ask how spins being up/down influence their joint probability? e.g. interactions amongst spins in a network

Start with 1 spin:
$$I_i^m = \ln\left(\frac{p(G_i = 1 \mid \underline{G} = 0)}{p(G_i = 0 \mid \underline{G} = 0)}\right)$$

'odds ratio': What is the likelihood of spin i being 1 vs 0

$$I_{i,j}^m = \ln\boxed{\left(\frac{p(G_{ij} = (1,1) \mid \underline{G} = 0)}{p(G_{ij} = (0,1) \mid \underline{G} = 0)}\right)} - \ln\boxed{\left(\frac{p(G_{ij} = (1,0) \mid \underline{G} = 0)}{p(G_{ij} = (0,0) \mid \underline{G} = 0)}\right)}$$

'odd ratio' of spin i with spin j being 1

'odd ratio' of spin i with spin j being 0

'generalised odds ratio': Does the likelihood of spin i being 1 increase/decrease depending on whether spin j is 1/0.

Multiplicative formulation of interactions

Alternatively: Do not wish to specify an outcome, instead ask how spins being up/down influence their joint probability? e.g. interactions amongst spins in a network

Start with 1 spin: $$I_i^m = \ln\left(\frac{p(G_i = 1 \mid \underline{G} = 0)}{p(G_i = 0 \mid \underline{G} = 0)}\right)$$

'odds ratio': What is the likelihood of spin i being 1 vs 0

$$I_{i,j}^m = \ln\left(\frac{p(G_{ij} = (1,1) \mid \underline{G} = 0)}{p(G_{ij} = (0,1) \mid \underline{G} = 0)}\right) - \ln\left(\frac{p(G_{ij} = (1,0) \mid \underline{G} = 0)}{p(G_{ij} = (0,0) \mid \underline{G} = 0)}\right)$$

If two spins are independent: $$p(G_i, G_j) = p(G_i)p(G_j)$$

There is no interaction: $$I_{i,j}^m = 0$$

Multiplicative formulation of interactions

Alternatively: Do not wish to specify an outcome, instead ask how spins being up/down influence their joint probability? e.g. interactions amongst spins in a network

Start with 1 spin: $$I_i^m = \ln\left(\frac{p(G_i = 1 \mid \underline{G} = 0)}{p(G_i = 0 \mid \underline{G} = 0)}\right)$$

'odds ratio': What is the likelihood of spin i being 1 vs 0

Higher-order interactions:

$$I_{i_1,\ldots,i_n}^m = \prod_{k=0}^{n}\left(\prod_{J\subset I:\,\ell(J)=k} p\big(G_I = e_J^{(n)} \mid \underline{G} = 0\big)^{(-1)^{n-k}}\right)$$

Where $e_J^{(\ell(I))} = (e_{i_1}, \ldots, e_{i_{\ell(I)}})$ is a tuple of elements taking values 0 or 1 (k)

Beentjes & Khamseh, Physical Review E (2020)

Challenge: Large number of dependent variables

Estimating intricate interaction structure amongst many genes

Certain approximation no longer possible: $p(G_i, G_j) \neq p(G_i)p(G_j)$

Number of variables >> data, (and high temperatures)

$$I^m_{i,j} = \ln\left(\frac{p(G_{ij} = (1,1) \mid \underline{G} = 0)}{p(G_{ij} = (0,1) \mid \underline{G} = 0)} \frac{p(G_{ij} = (0,0) \mid \underline{G} = 0)}{p(G_{ij} = (1,0) \mid \underline{G} = 0)}\right)$$

Estimate conditional dependencies directly from data, using efficient causal discovery algorithms (e.g. PC, Score-based MCMC)

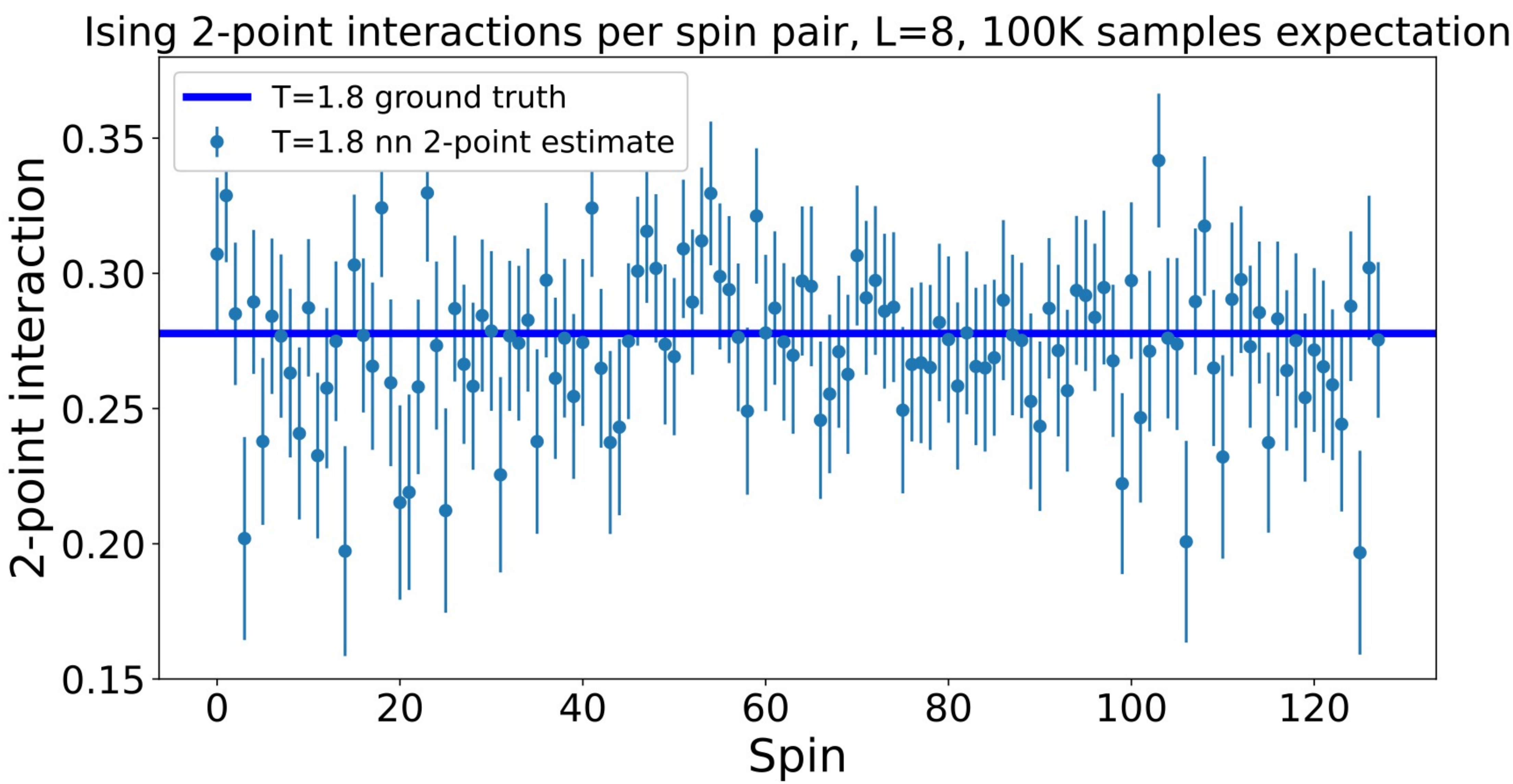
Nothing comes for free! These come with their own assumptions/bias
Keep in mind to be conservative.

Model-independent interaction estimator on 2D Ising

Back to Ising …

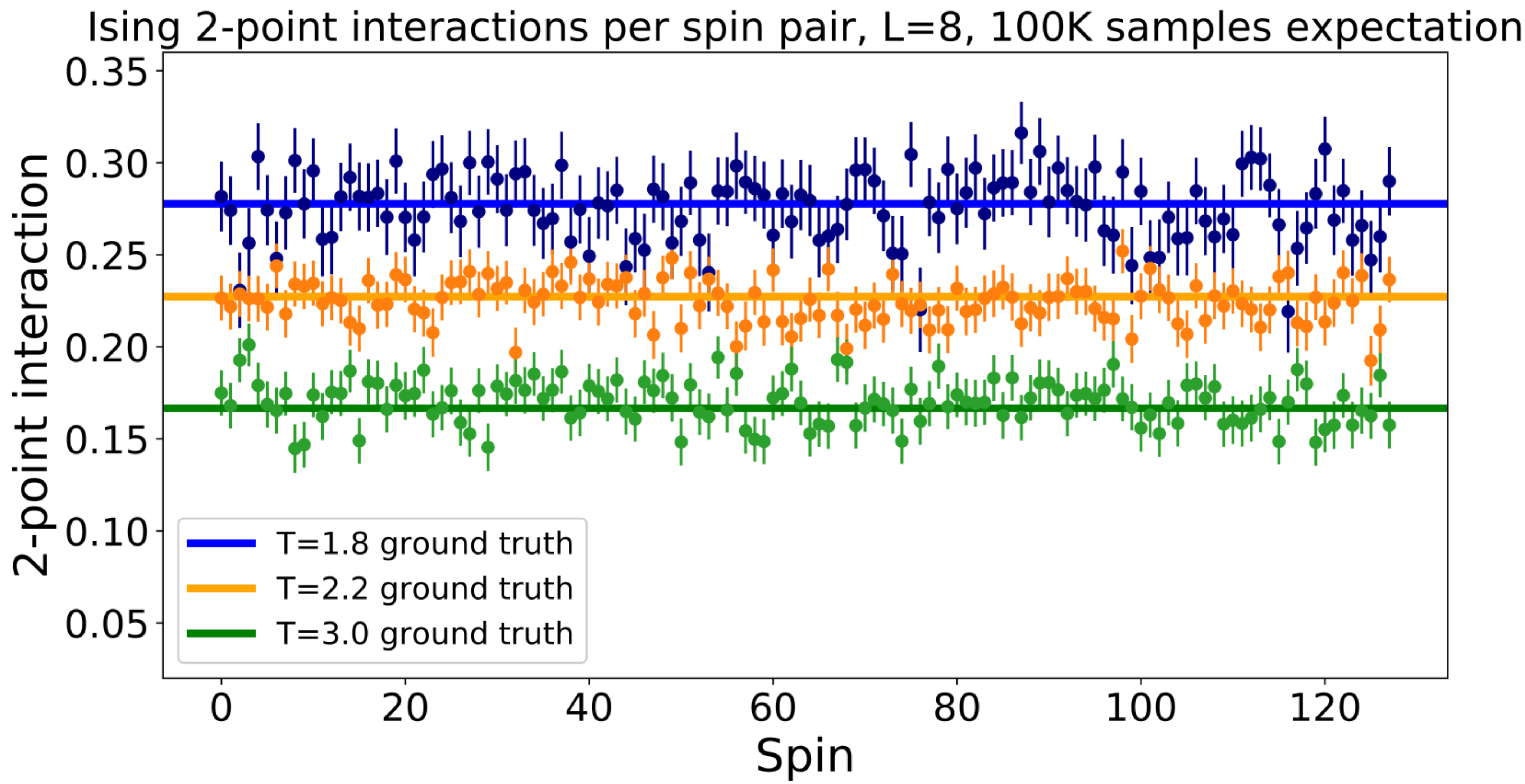
$$p_D(s) = \frac{1}{Z(J,h)}\, e^{-H_{J,h}(s)}$$

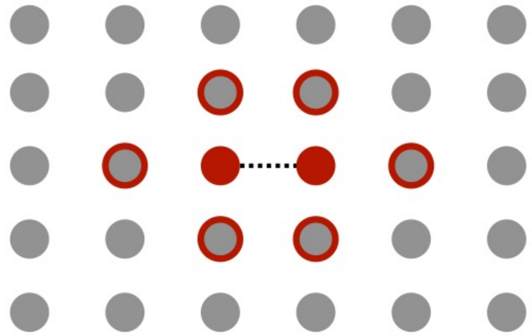
$$H_{J,h} = -\sum J_{ij} s_i s_j - \sum h_i s_i \quad , \quad Z(J,h) = \sum e^{-H_{J,h}(s)}$$



Interactions: Numerical Results using CI

Statistical physics system: Ising model in 2D, 64 variables





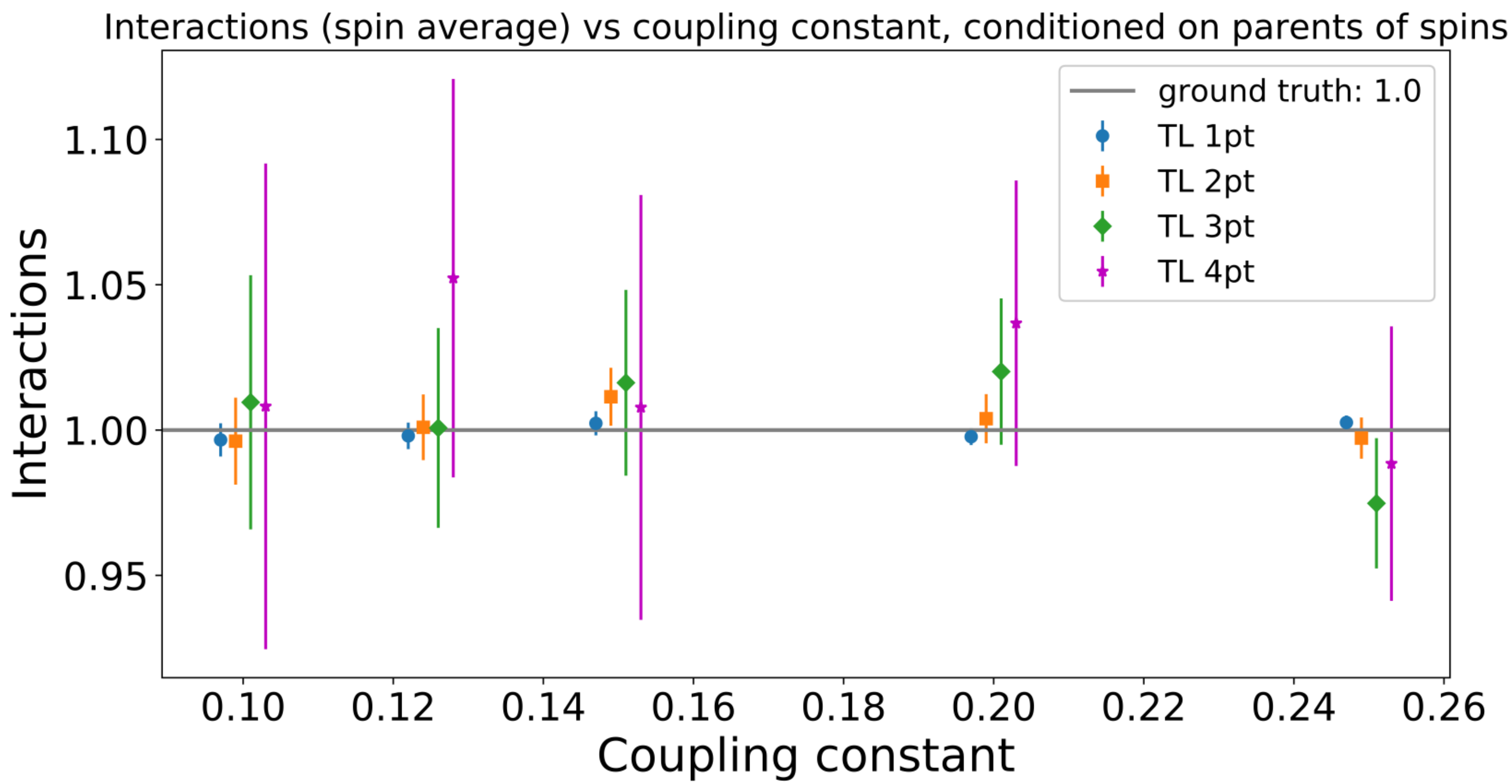
$$H(s) = -\sum_{\langle i,j \rangle} J_{ij} s_i s_j$$

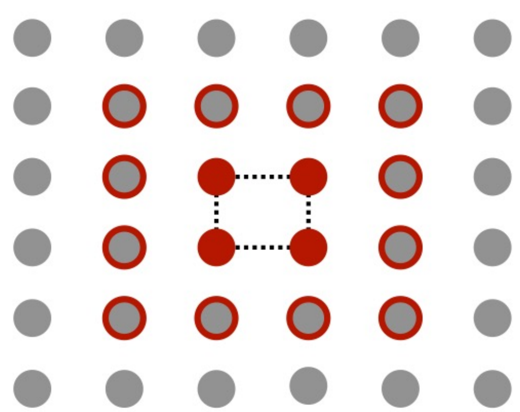
$$p(s) = \frac{1}{Z} e^{-H(s)}$$

Beentjes & Khamseh, Physical Review E (2020)

Interactions: Numerical Results

Statistical physics system: Ising-type model, 4-point interactions



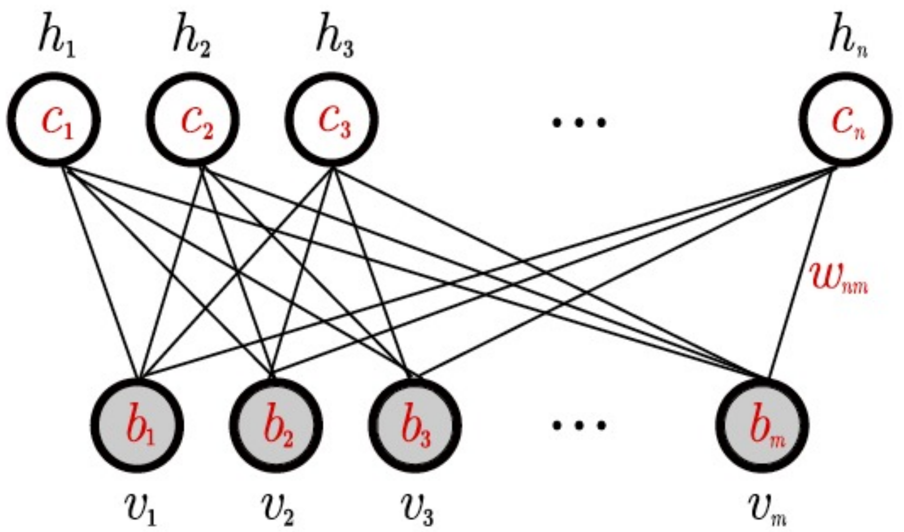


Beentjes & Khamseh, Physical Review E (2020)

Recall: Restricted Boltzmann Machine

$$E_\theta(\mathbf{v}, \mathbf{h}) = -\sum_{i=1}^{n}\sum_{j=1}^{m} w_{ij} h_i v_j - \sum_{i=1}^{n} c_i h_i - \sum_{j=1}^{m} b_j v_j$$

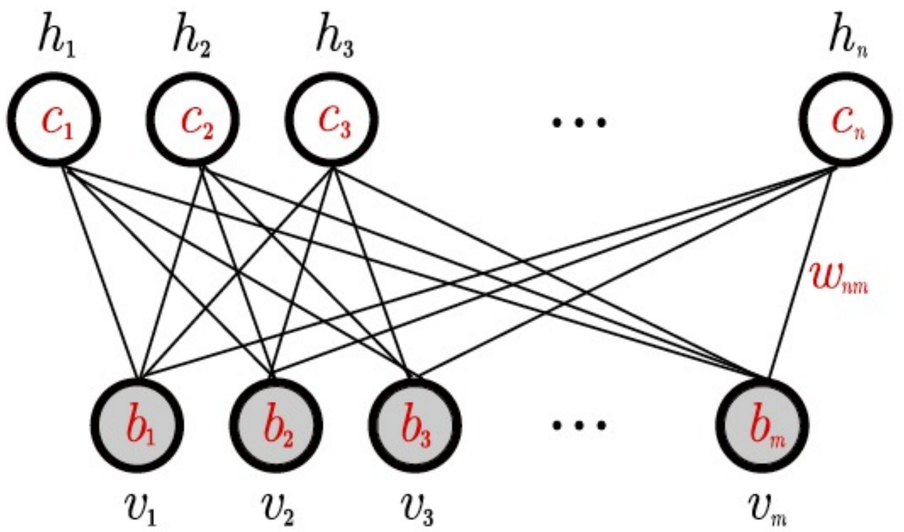
$$\boxed{p_{\text{RBM}}(\mathbf{v}, \mathbf{h}|\theta) = \frac{1}{Z_{\text{RBM}}} e^{-E_\theta(\mathbf{v},\mathbf{h})}}$$



Recall: Restricted Boltzmann Machine

$$E_\theta(\mathbf{v}, \mathbf{h}) = -\sum_{i=1}^{n} \sum_{j=1}^{m} w_{ij} h_i v_j - \sum_{i=1}^{n} c_i h_i - \sum_{j=1}^{m} b_j v_j$$

$$\boxed{p_{\text{RBM}}(\mathbf{v}, \mathbf{h}|\theta) = \frac{1}{Z_{\text{RBM}}} e^{-E_\theta(\mathbf{v},\mathbf{h})}}$$



Marginal: $$p(\mathbf{v}|\theta) = \frac{1}{\mathcal{Z}(\theta)} \prod_{j=1}^{m} \left(e^{b_j v_j}\right) \prod_{i=1}^{n} \left(1 + e^{c_i + \sum_{j=1}^{m} w_{ij} v_j}\right)$$

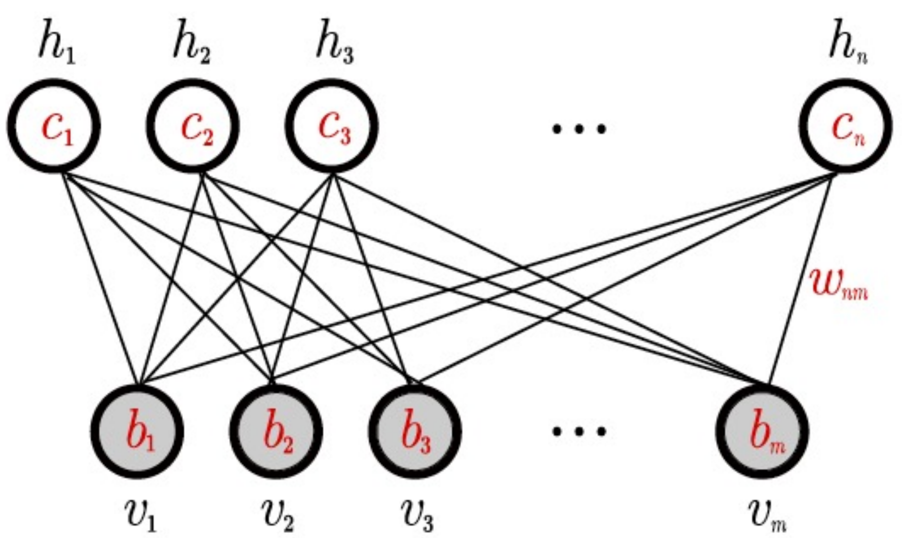
Asymptotic expansion, resummation, …

Analytical closed-form expression for n-point interactions, e.g. 2-point:

$$J_{j_1,j_2} \propto \ln \prod_{i=1}^{n} \frac{(1 + e^{c_i + w_{ij_1} + w_{ij_2}})(1 + e^{c_i})}{(1 + e^{c_i + w_{ij_1}})(1 + e^{c_i + w_{ij_2}})}$$

Restricted Boltzmann Machine

$$E_\theta(\mathbf{v},\mathbf{h}) = -\sum_{i=1}^{n}\sum_{j=1}^{m} w_{ij} h_i v_j - \sum_{i=1}^{n} c_i h_i - \sum_{j=1}^{m} b_j v_j$$



$$\boxed{p_{\mathrm{RBM}}(\mathbf{v},\mathbf{h}|\theta) = \frac{1}{Z_{\mathrm{RBM}}} e^{-E_\theta(\mathbf{v},\mathbf{h})}}$$

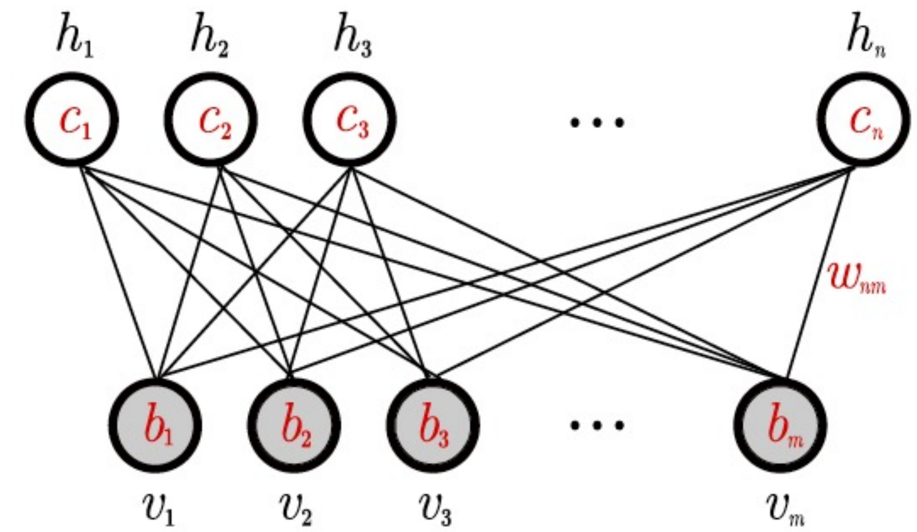
Marginal: $$p(\mathbf{v}|\theta) = \frac{1}{\mathcal{Z}(\theta)} \prod_{j=1}^{m} \left(e^{b_j v_j}\right) \prod_{i=1}^{n} \left(1 + e^{c_i + \sum_{j=1}^{m} w_{ij} v_j}\right)$$

Instead, use the TL formulation to directly read-off the coupling!!

$$I^m_{j_1,j_2} = \frac{p(v_{j_1 j_2} = (1,1), \underline{v} = 0)\, p(v_{j_1 j_2} = (0,0), \underline{v} = 0)}{p(v_{j_1 j_2} = (1,0), \underline{v} = 0)\, p(v_{j_1 j_2} = (0,1), \underline{v} = 0)} = \prod_{i=1}^{n} \frac{(1 + e^{c_i + w_{ij_1} + w_{ij_2}})(1 + e^{c_i})}{(1 + e^{c_i + w_{ij_1}})(1 + e^{c_i + w_{ij_2}})}$$

Restricted Boltzmann Machine

$$E_\theta(\mathbf{v},\mathbf{h}) = -\sum_{i=1}^{n}\sum_{j=1}^{m} w_{ij}h_i v_j - \sum_{i=1}^{n} c_i h_i - \sum_{j=1}^{m} b_j v_j$$



$$\boxed{p_{\text{RBM}}(\mathbf{v},\mathbf{h}|\theta) = \frac{1}{Z_{\text{RBM}}} e^{-E_\theta(\mathbf{v},\mathbf{h})}}$$

Marginal: $$p(\mathbf{v}|\theta) = \frac{1}{\mathcal{Z}(\theta)} \prod_{j=1}^{m} \left(e^{b_j v_j}\right) \prod_{i=1}^{n} \left(1 + e^{c_i + \sum_{j=1}^{m} w_{ij} v_j}\right)$$

Instead, use the TL formulation to directly read-off the coupling!!

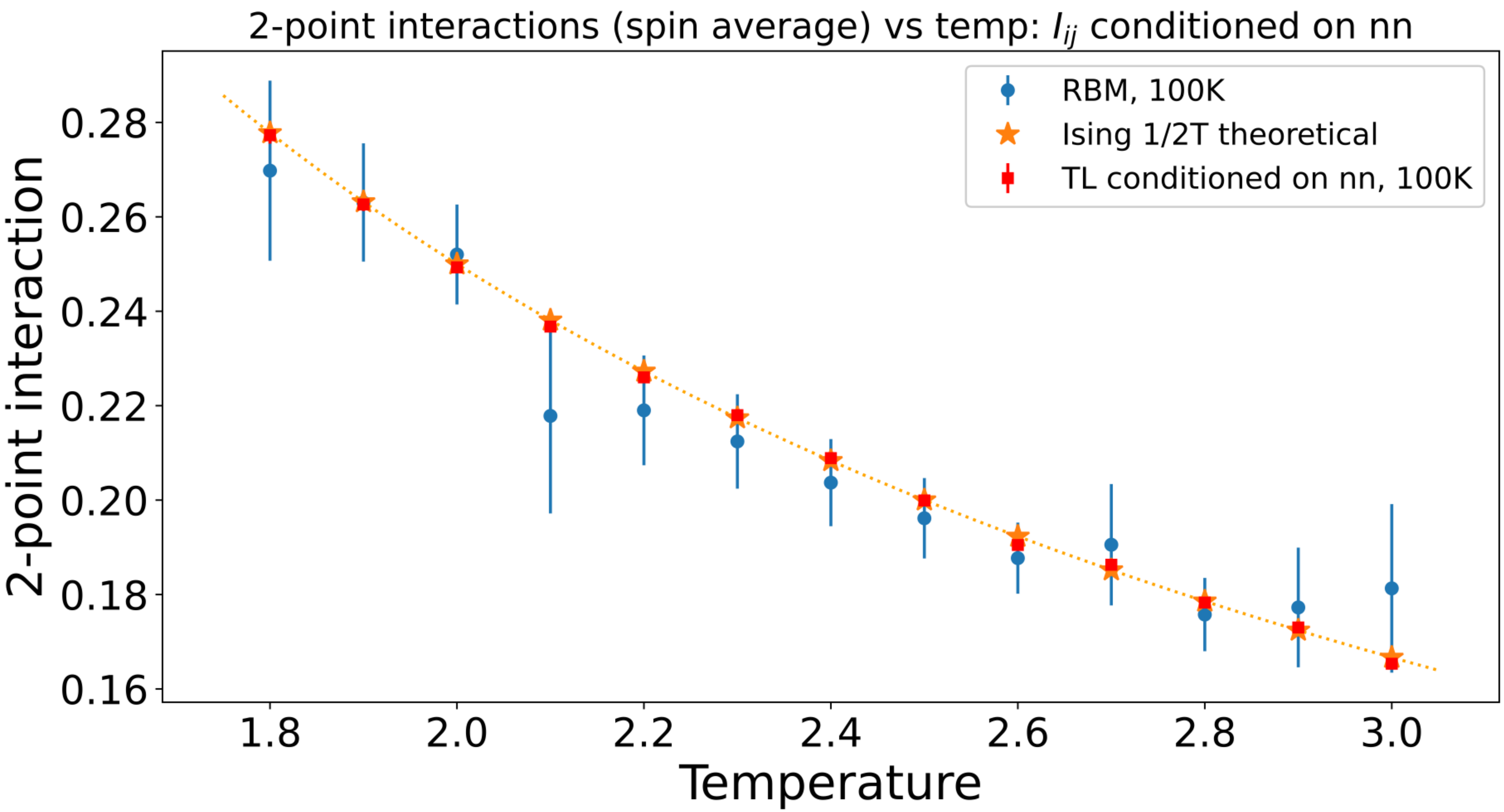
$$I^m_{j_1,j_2} = \frac{p(v_{j_1 j_2} = (1,1), \underline{v} = 0)}{p(v_{j_1 j_2} = (1,0), \underline{v} = 0)} \frac{p(v_{j_1 j_2} = (0,0), \underline{v} = 0)}{p(v_{j_1 j_2} = (0,1), \underline{v} = 0)} = \prod_{i=1}^{n} \frac{(1+e^{c_i + w_{ij_1} + w_{ij_2}})(1+e^{c_i})}{(1+e^{c_i+w_{ij_1}})(1+e^{c_i+w_{ij_2}})}$$

No asymptotic expansion and resummation required …

Applies to other energy based models

Improving estimation via conditional independence

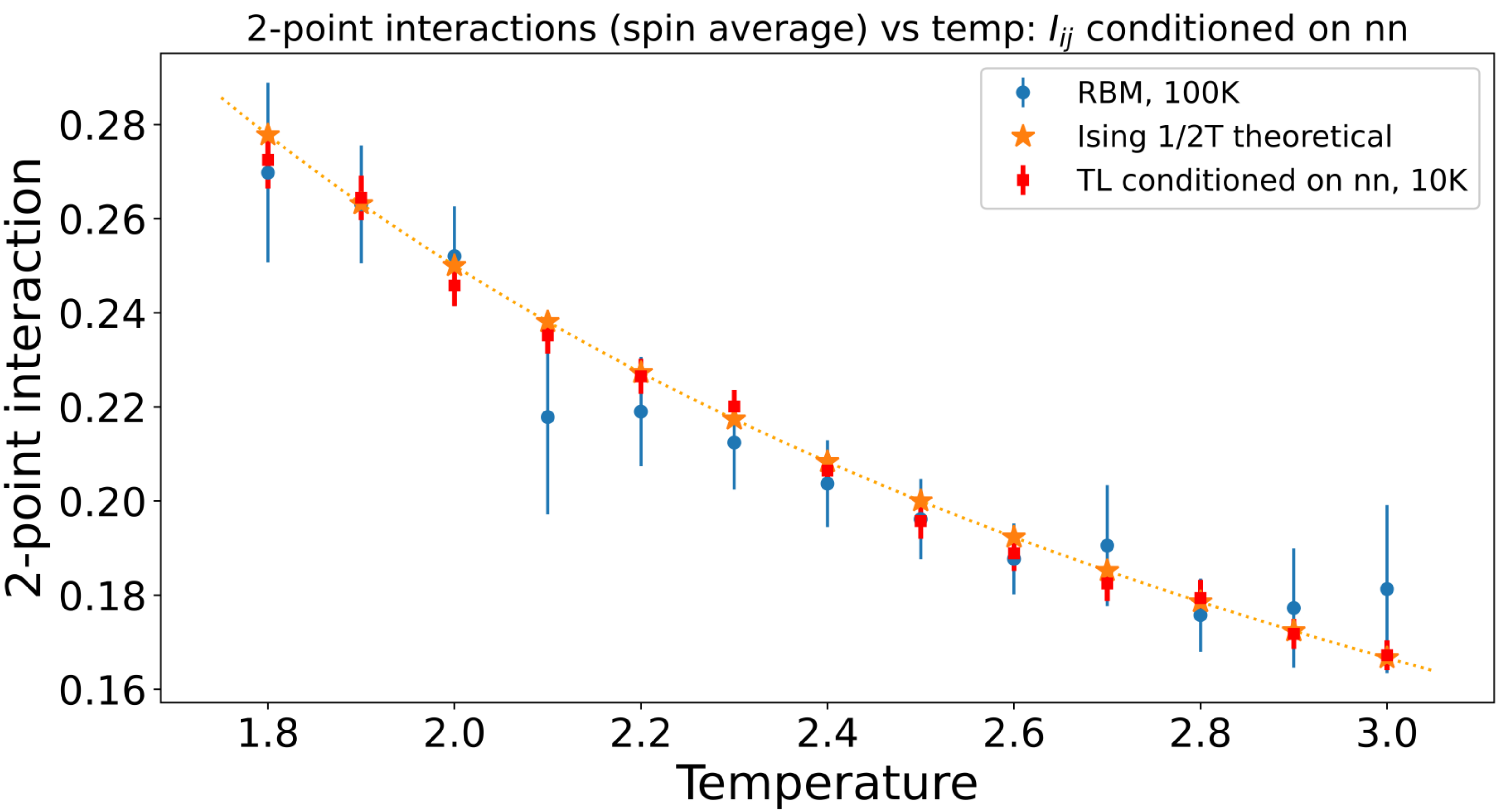
Conditioning on parent spins to isolate pairs from the rest of the system (Markovian). Run time: Few seconds per temperature.



100K samples

Improving estimation via conditional independence

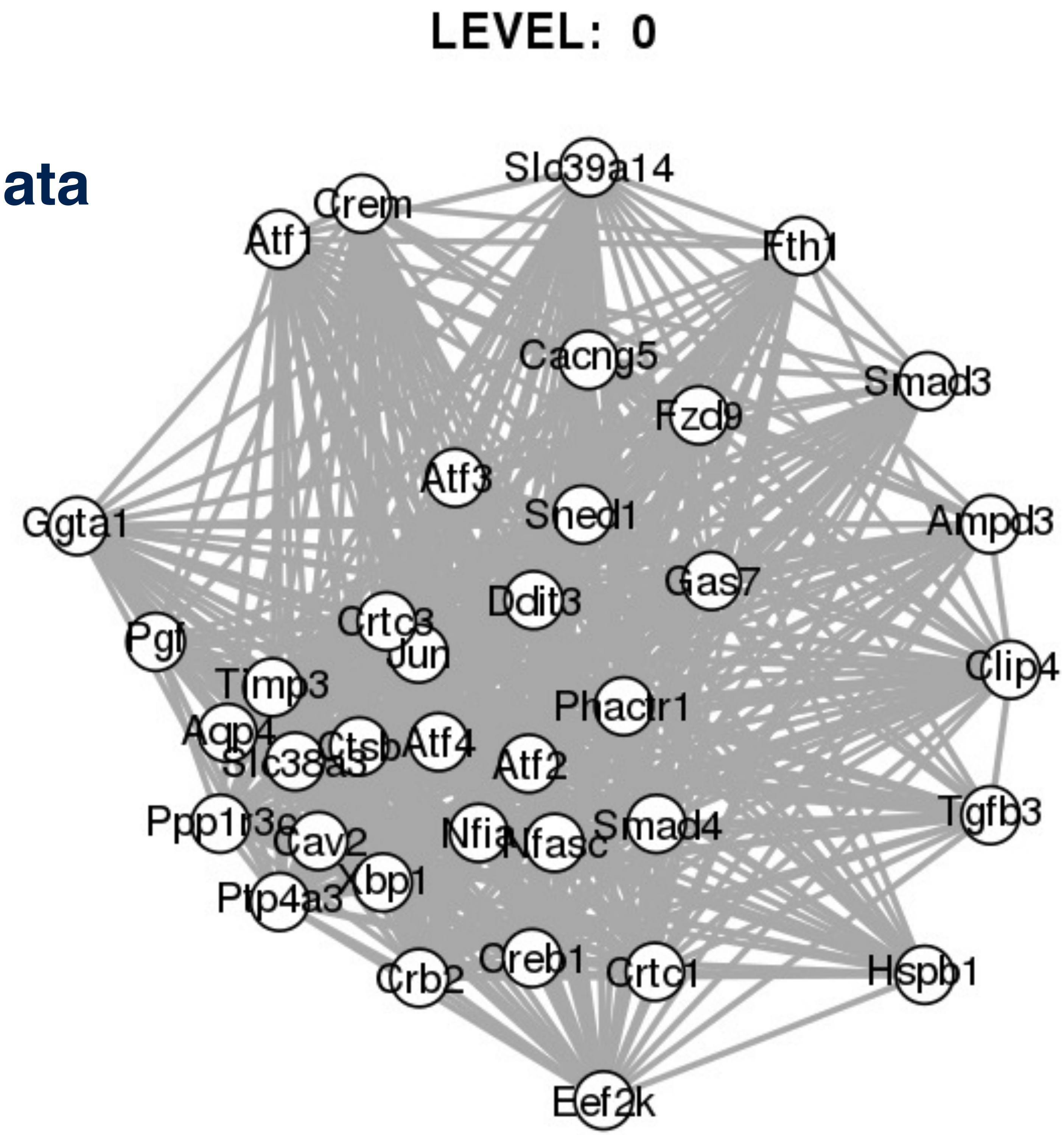
Conditioning on parent spins to isolate pairs from the rest of the system (Markovian). Run time: Few seconds per temperature.



10K samples

Gene Networks: Independence and interactions

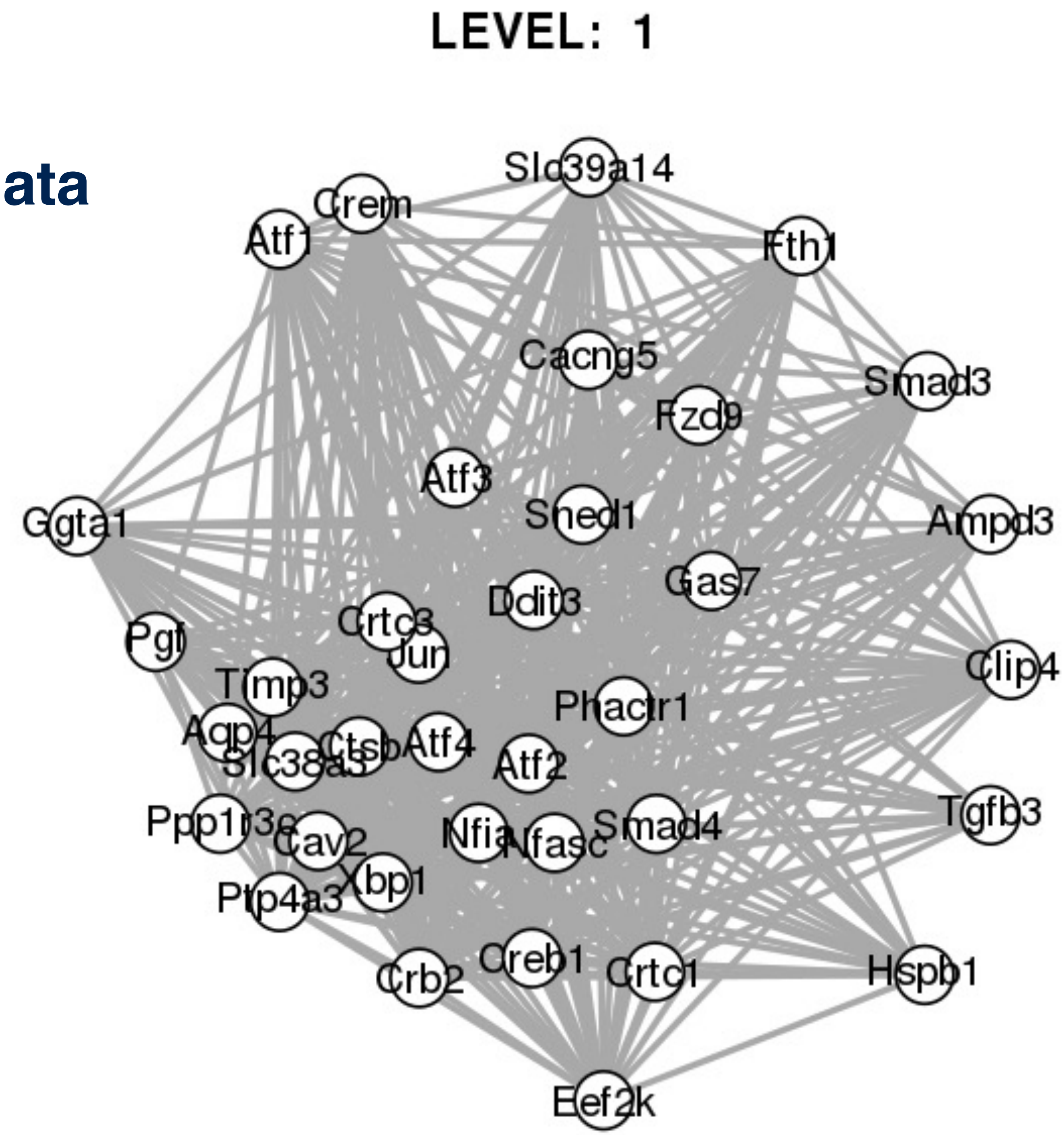
10X single-cell
1M mouse brain
developmental data



Plot by Abel Jansma (PhD student)

Causal Discovery in Gene Networks: An example

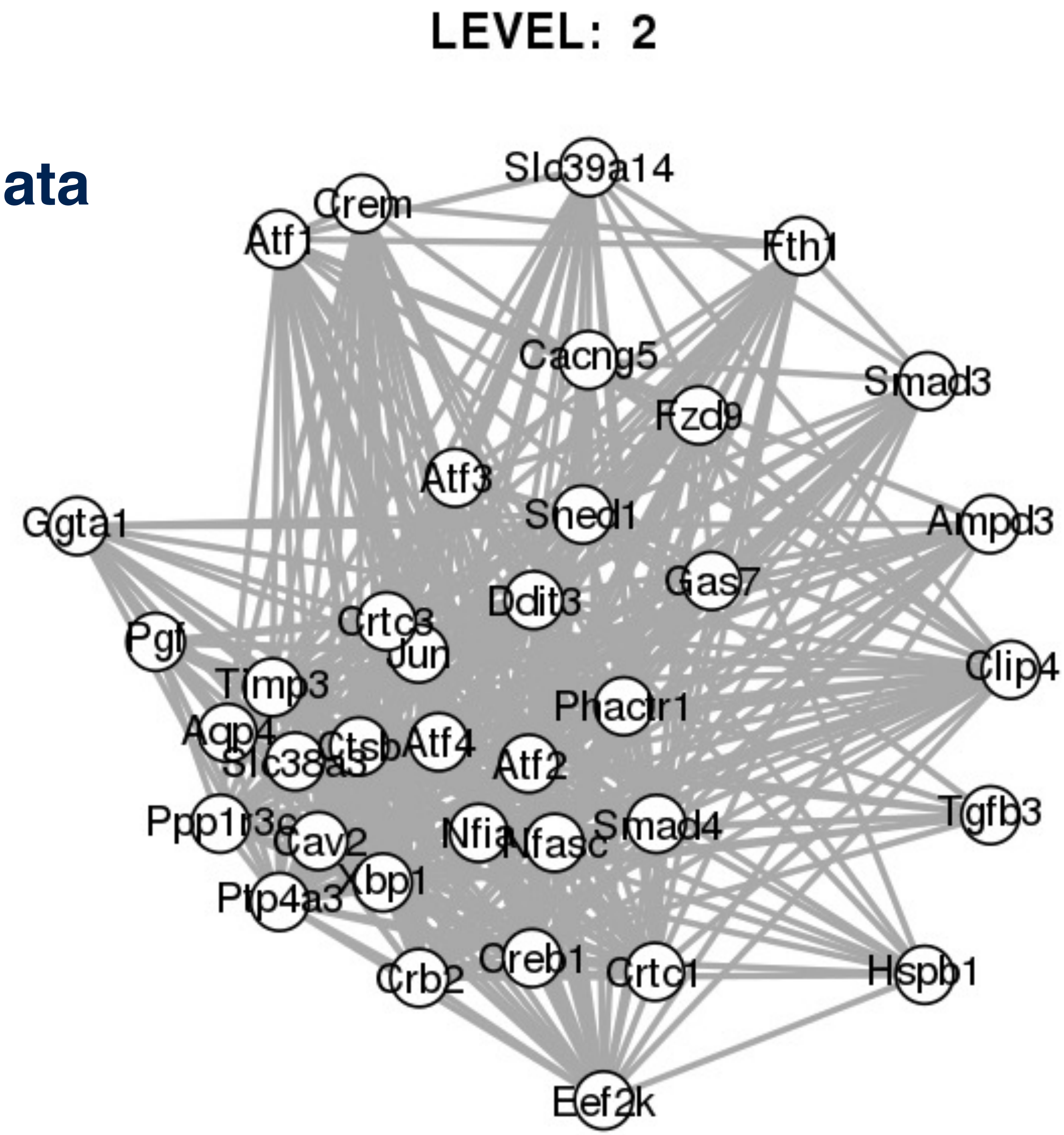
10X single-cell
1M mouse brain
developmental data



Plot by Abel Jansma (PhD student)

Causal Discovery in Gene Networks: An example

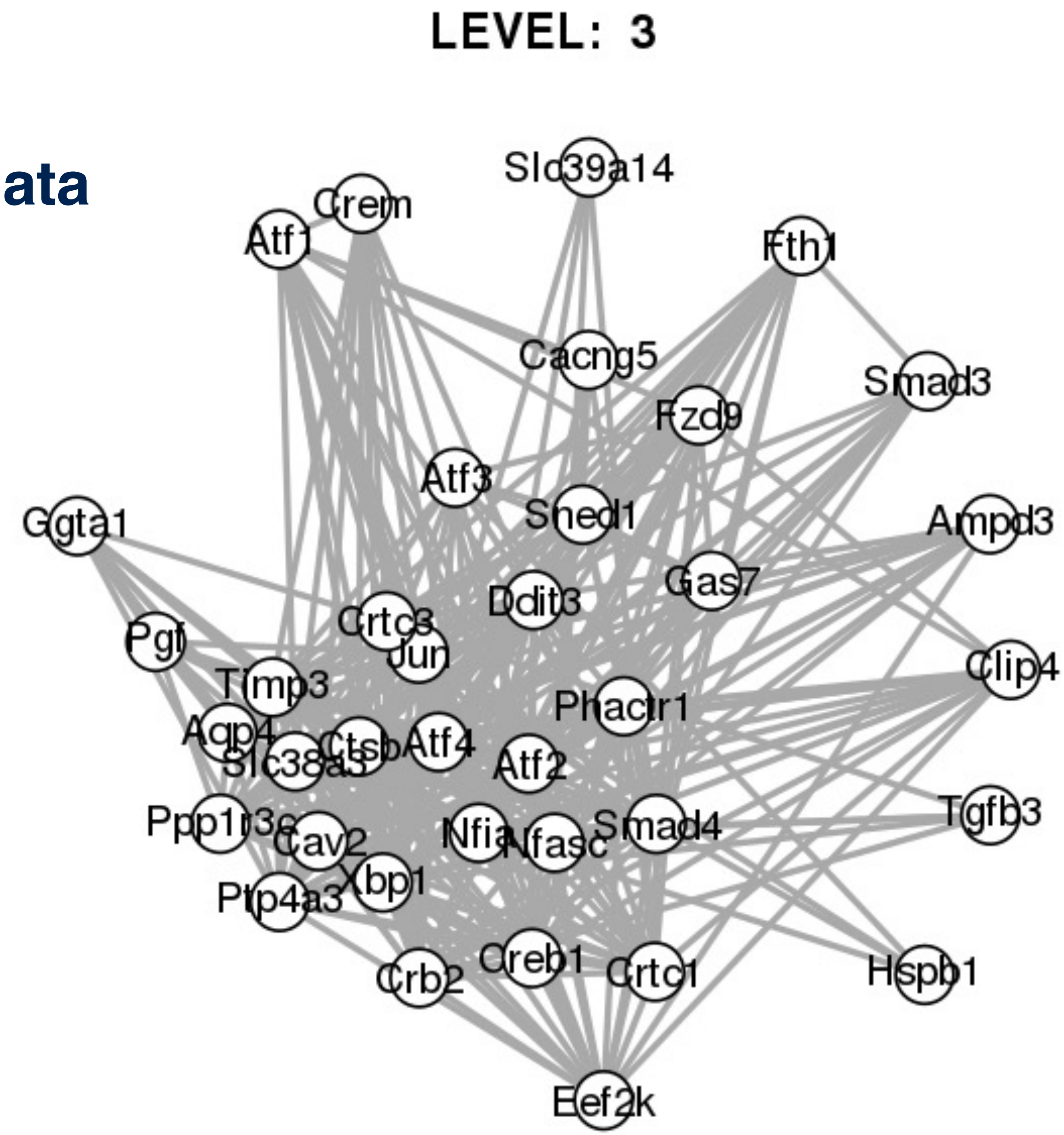
10X single-cell
1M mouse brain
developmental data



Plot by Abel Jansma (PhD student)

Causal Discovery in Gene Networks: An example

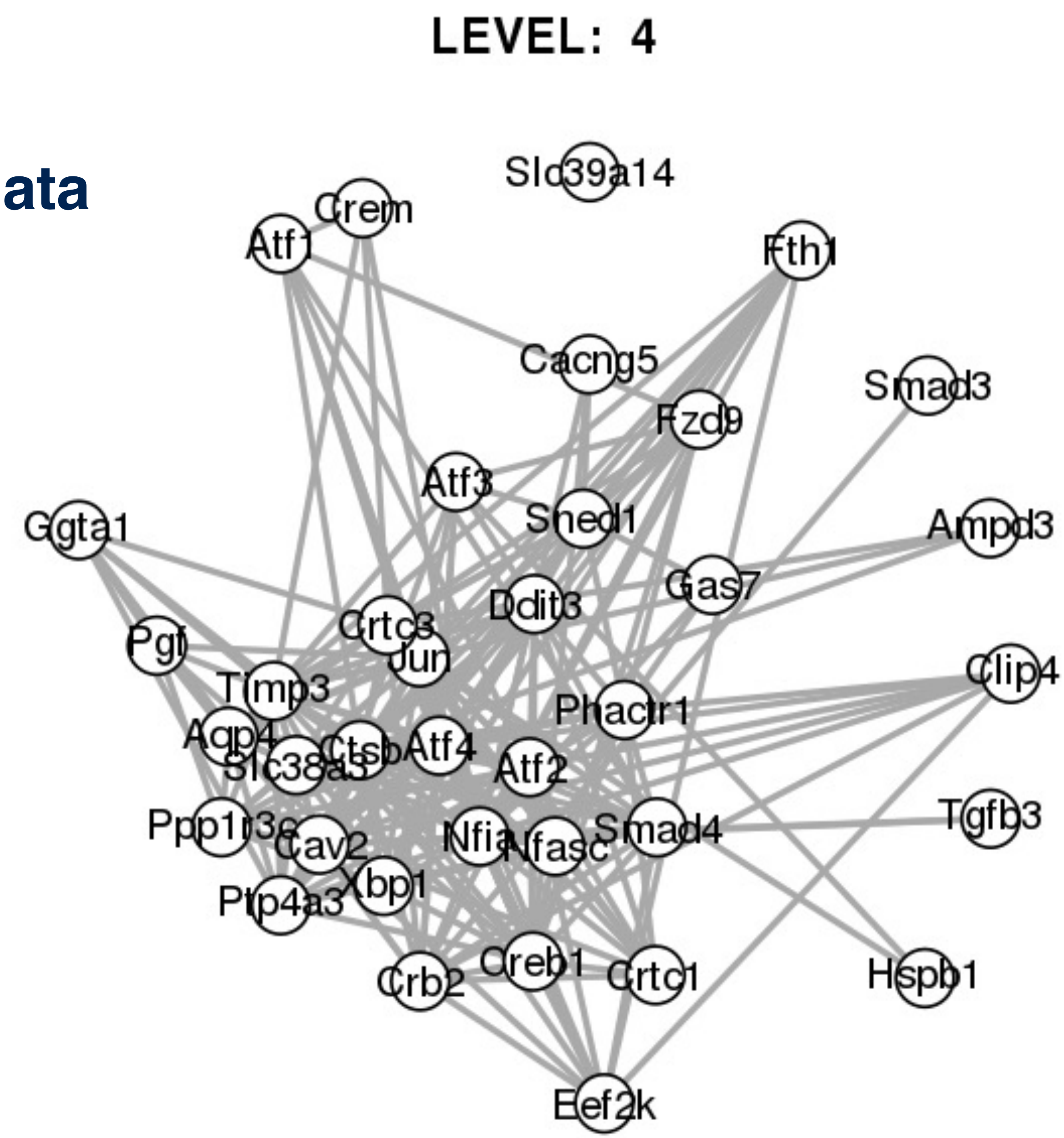
10X single-cell
1M mouse brain
developmental data



Plot by Abel Jansma (PhD student)

Causal Discovery in Gene Networks: An example

10X single-cell
1M mouse brain
developmental data



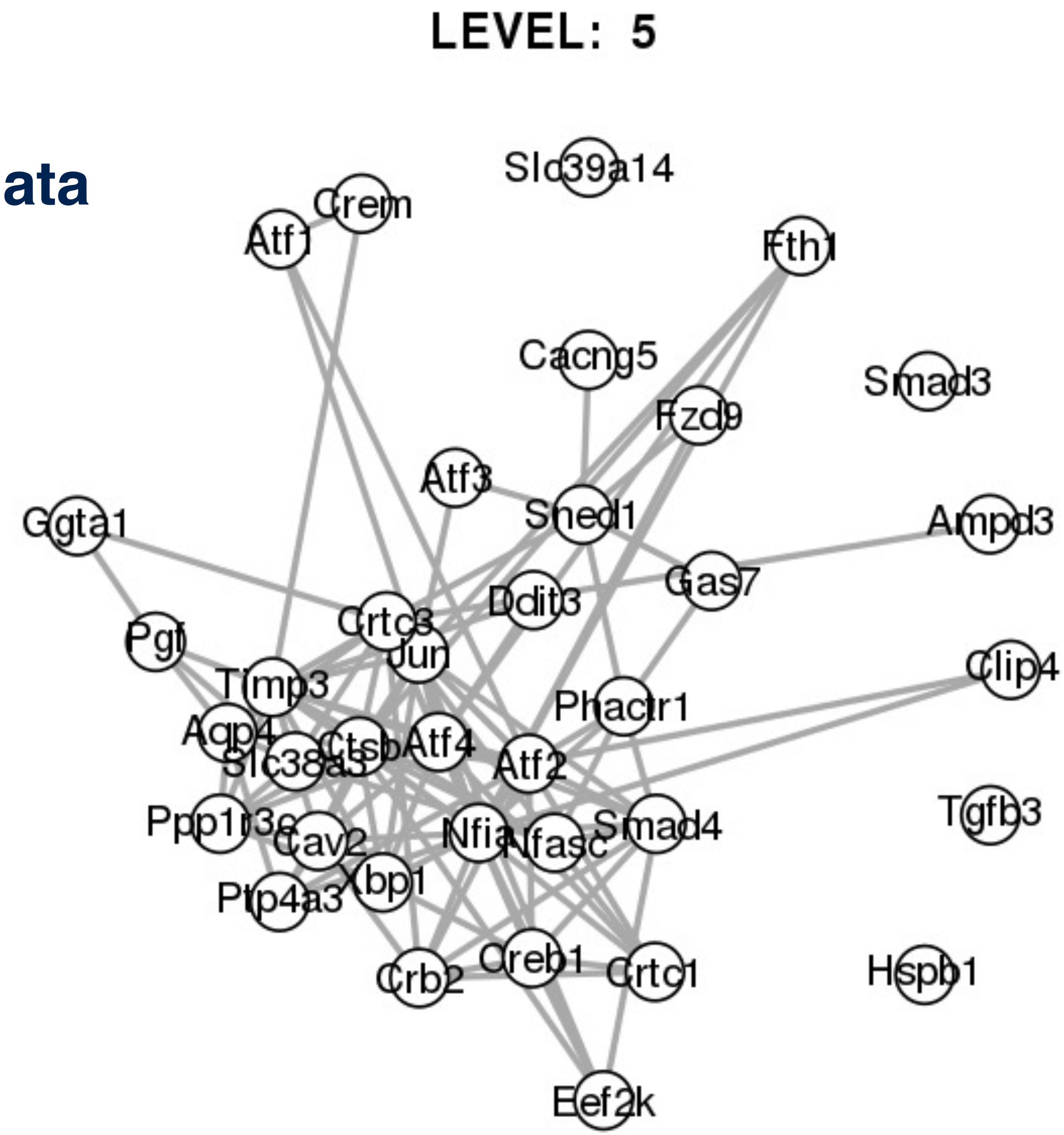
Plot by Abel Jansma (PhD student)

Causal Discovery in Gene Networks: An example

10X single-cell

1M mouse brain

developmental data



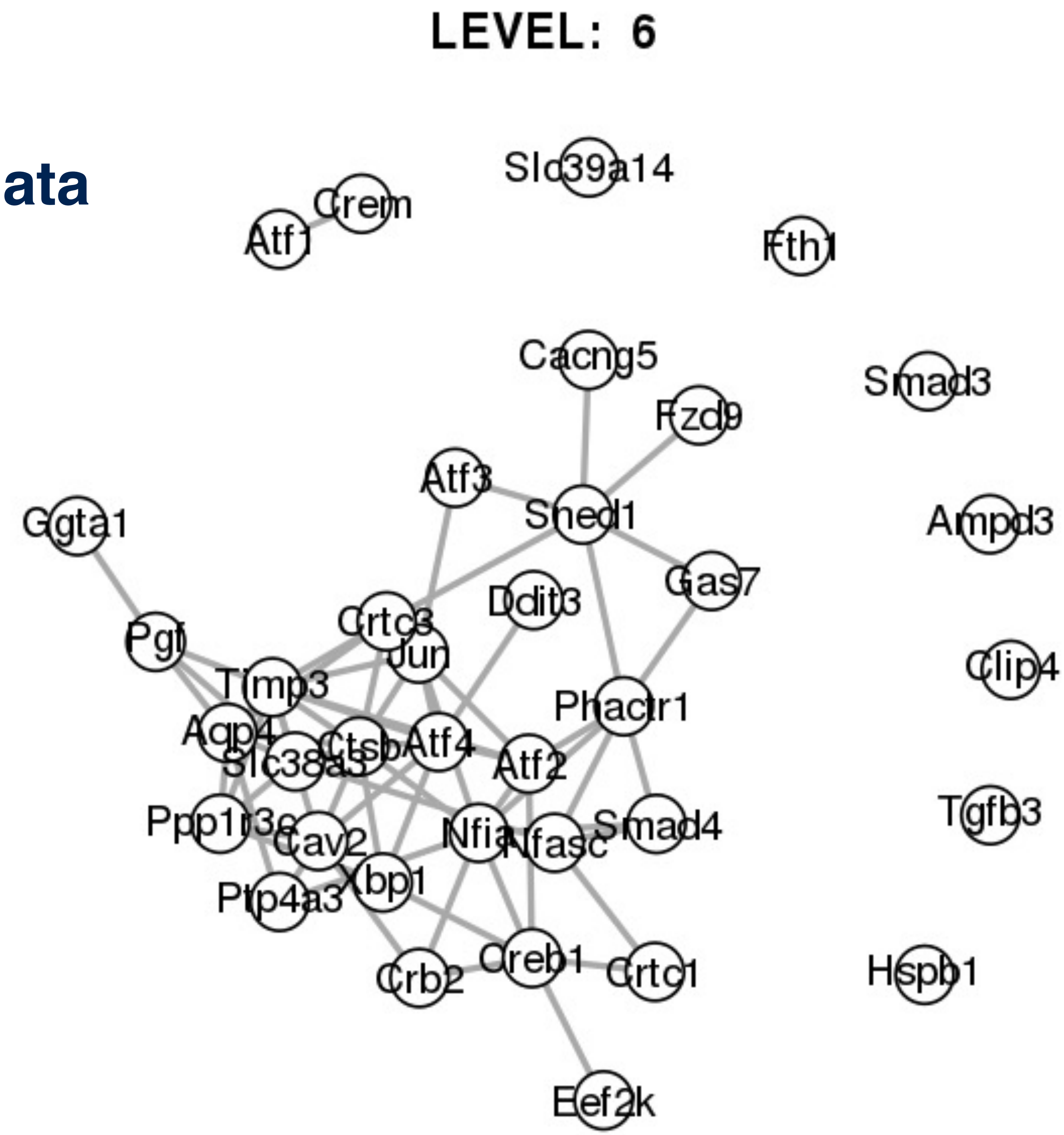
Plot by Abel Jansma (PhD student)

Causal Discovery in Gene Networks: An example

10X single-cell

1M mouse brain

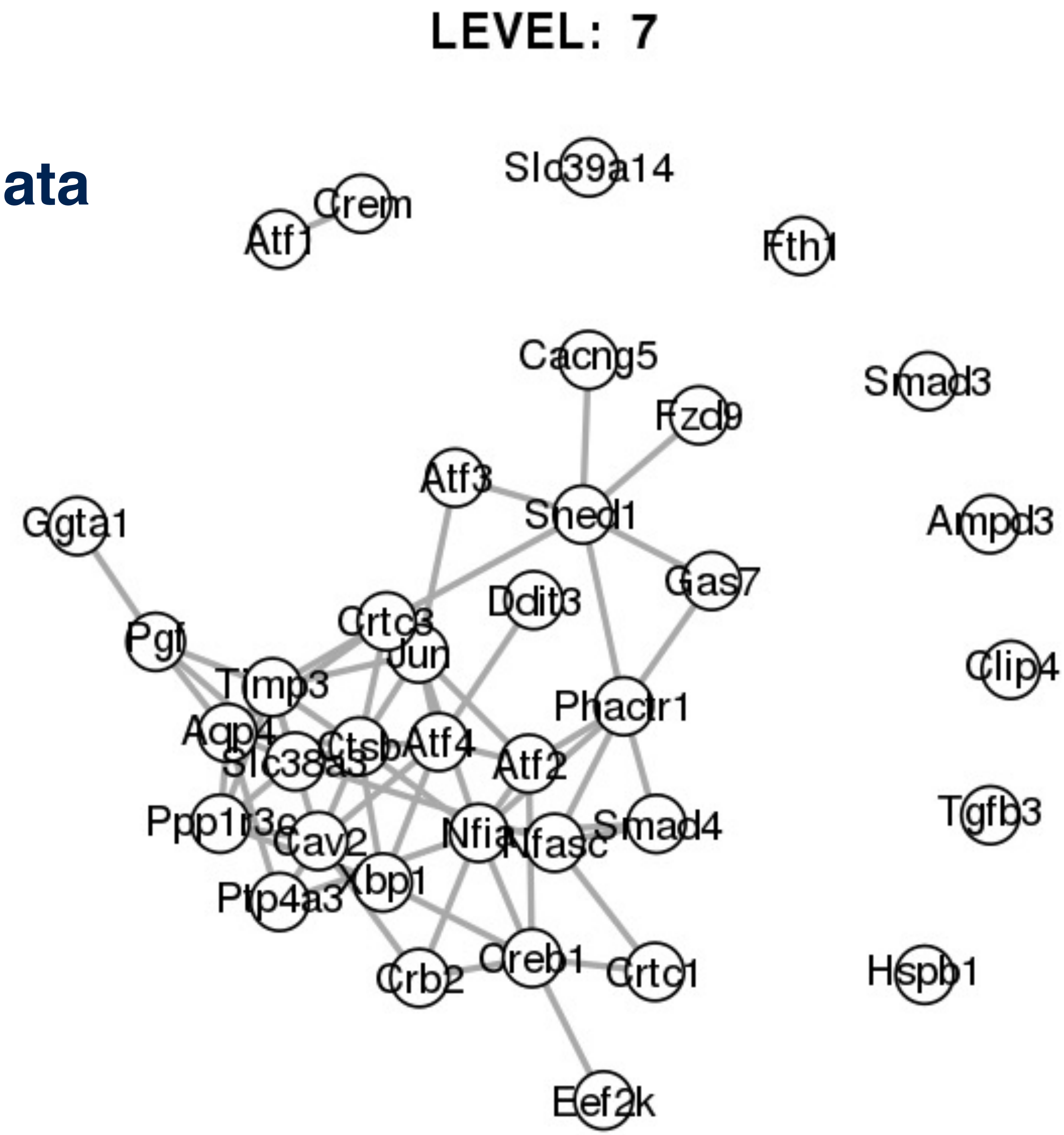
developmental data



Plot by Abel Jansma (PhD student)

Causal Discovery in Gene Networks: An example

10X single-cell
1M mouse brain
developmental data



Plot by Abel Jansma (PhD student)

Conclusions

We have provided a non-parametric solution to the inverse problem of estimating n-point interactions for binary (and categorical) variables

Fully model-independent and unbiased
(no specific probability distribution is assumed)

Can extract interaction for any parametric model
(e.g. energy based neural networks)

Estimators consist of only computing expectation values over the data,
run time: few minutes on a local machine

Maximal use of data by targeting the quantity of interest directly

Higher-order interactions in statistical physics, machine learning (and biomedicine)

Ava Khamseh
Schools of Informatics and Physics
Institute of Genetics and Cancer

28 Sep 2021