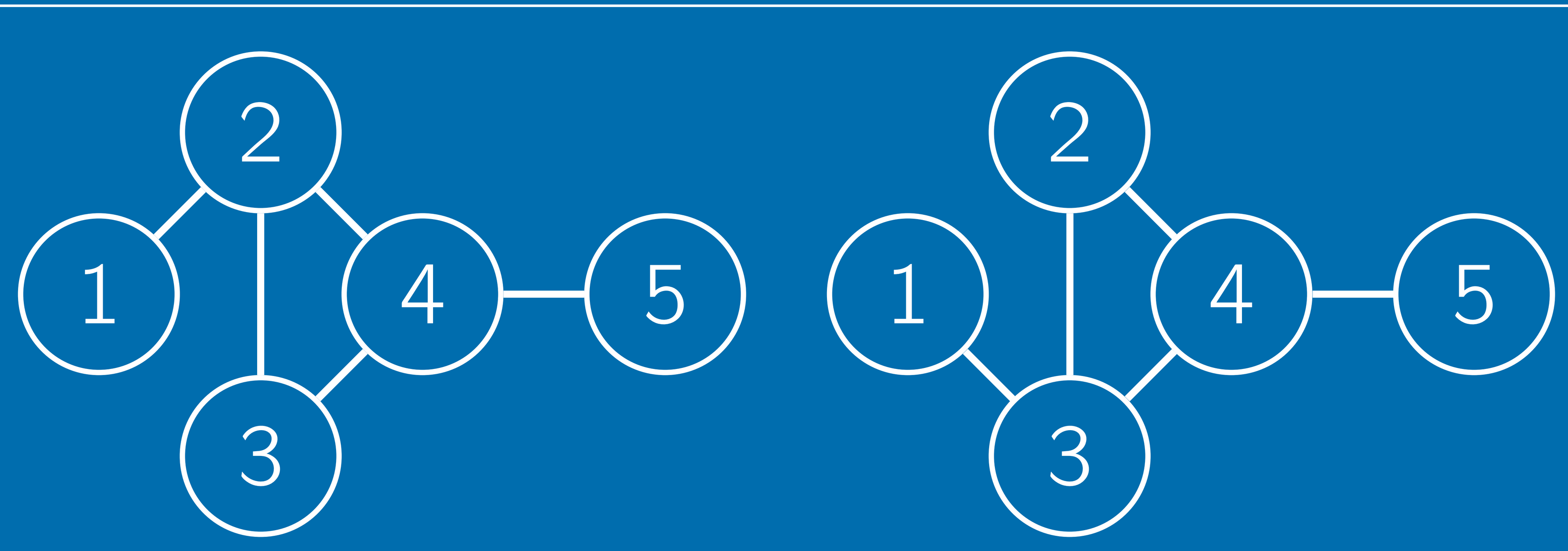
Compute a **family** of **divergences** between discrete **graphical models** using **belief propagation.**



$$P = \frac{P_{1,2} \cdot P_{2,3,4} \cdot P_{4,5}}{P_2 \cdot P_4} = P_{2,3,4} \cdot P_{1|2} \cdot P_{5|4} \qquad Q = \frac{Q_{1,3} \cdot Q_{2,3,4} \cdot Q_{4,5}}{Q_3 \cdot Q_4} = Q_{2,3,4} \cdot Q_{1|3} \cdot Q_{5|4}$$

$$D_{AB}^{(\alpha,\beta)}(P,Q) \in \mathcal{O}\left(n^2 \omega(\mathcal{H}) \cdot 2^{\omega(\mathcal{H})+1}\right)$$

Loong Kuan Lee [1]
Nico Piatkowski [2]
François Petitjean [1]
Geoffrey I. Webb [1]

[1]Monash University
[2]Fraunhofer IAIS

Scan with a phone, or go to **https://lklee.dev/pubs/2023-aaai** , to download the **full paper** and other **supplementary materials**

Motivation

Compute the divergence between 2 discrete distributions. For instance take the KL divergence:

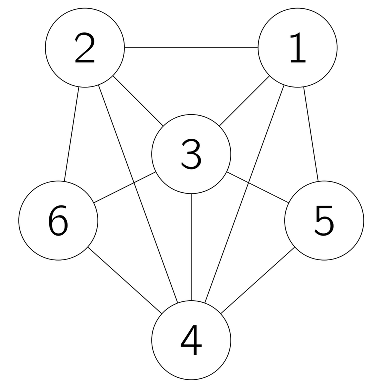
$$D_{\mathrm{KL}}(P,Q) = \sum_{x \in \mathcal{X}} P(x) \log \frac{P(x)}{Q(x)}$$
$$\mathcal{O}(\mathcal{X}) \in \mathcal{O}(2^n)$$

However, complexity is exponential w.r.t. the number of variables, and therefore intractable.

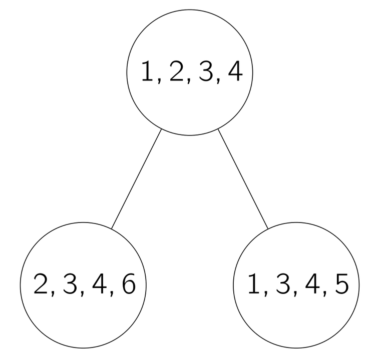
Background

Decomposable Models

Markov networks with a choral graph structure. They have a direct clique tree representation:



Chordal graph



Clique tree

and a closed form expression of its distribution:

$$P = \frac{P_{1,2,3,4} \cdot P_{2,3,4,6} \cdot P_{1,3,4,5}}{P_{2,3,4} \cdot P_{1,3,4}} = P_{1,2,3,4} \cdot P_{6|2,3,4} \cdot P_{5|1,3,4}$$

which more generally has the form:

$$P = \frac{\prod_{\mathcal{C} \in \mathcal{C}} P_{\mathcal{C}}}{\prod_{\mathcal{S} \in \mathcal{S}} P_{\mathcal{S}}} = \prod_{\mathcal{C} \in \mathcal{C}} P_{\mathcal{C}}^{\mathcal{T}}$$

$\alpha\beta$-divergence

A family of divergences that can express other divergences such as the KL-divergence. We showed that the $\alpha\beta$-divergence can be expressed as follows:

$$D_{\mathrm{AB}}^{(\alpha,\beta)}(P,Q) = \begin{cases} \sum_{x\in\mathcal{X}} \frac{1}{2}(\log P(x) - \log Q(x))^2 & \alpha,\beta = 0 \\ \mathcal{F}^{(1)}(P,Q) + \ldots + \mathcal{F}^{(k)}(P,Q) & \text{otherwise} \end{cases}$$

$$\mathcal{F}[g,h,g^*,h^*](P,Q) = \sum_{x\in\mathcal{X}} \left[g[P](x)\right]\left[h[Q](x)\right] \log\left(\left[g^*[P](x)\right]\left[h^*[Q](x)\right]\right)$$

where for functional $f \in \{g, h, g^*, h^*\}$:

$$f\left[\prod_{\mathcal{C}\in\mathcal{C}} P_{\mathcal{C}}^{\mathcal{T}}\right](x) = \prod_{\mathcal{C}\in\mathcal{C}} f\left[P_{\mathcal{C}}^{\mathcal{T}}\right](x_s)$$

Computation when $\alpha, \beta = 0$

For $\omega = \max(\omega(\mathcal{G}_P), \omega(\mathcal{G}_Q))$:

$$D_{\mathrm{AB}}^{(0,0)}(P,Q) = \frac{1}{2}\left(\log \prod_{\mathcal{C}\in\mathcal{C}(\mathcal{G}_P)} P_{\mathcal{C}}^{\mathcal{T}} - \log \prod_{\mathcal{C}\in\mathcal{C}(\mathcal{G}_Q)} Q_{\mathcal{C}}^{\mathcal{T}}\right)^2$$
$$= \frac{1}{2}\left(\sum_{\mathcal{C}\in\mathcal{C}(\mathcal{G}_P)} \log P_{\mathcal{C}}^{\mathcal{T}} - \sum_{\mathcal{C}\in\mathcal{C}(\mathcal{G}_Q)} \log Q_{\mathcal{C}}^{\mathcal{T}}\right)^2$$
$$\in \mathcal{O}(n^2\omega 2^{\omega+1})$$

Computation when $\alpha \neq 0$ or $\beta \neq 0$

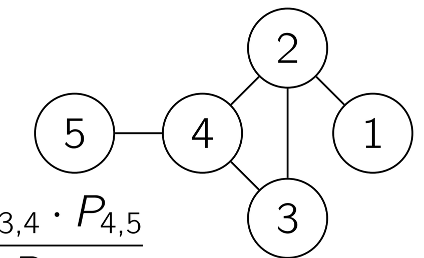
Substituting the distribution of decomposable models into $\mathcal{F}$:

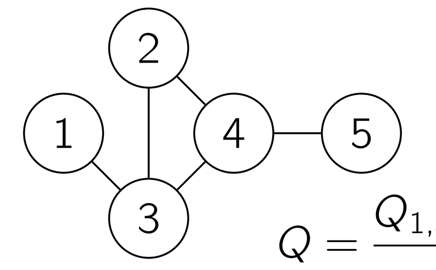
$$\mathcal{F}[g,h,g^*,h^*](P,Q) = \sum_{\mathcal{C}\in\mathcal{C}(\mathcal{G}_P)} \sum_{x_{\mathcal{C}}\in\mathcal{X}_{\mathcal{C}}} \log\left(g^*\left[P_{\mathcal{C}}^{\mathcal{T}}\right](x_{\mathcal{C}})\right) \mathrm{SP}_{\mathcal{C}}(x_{\mathcal{C}}) + \sum_{\mathcal{C}\in\mathcal{C}(\mathcal{G}_Q)} \sum_{x_{\mathcal{C}}\in\mathcal{X}_{\mathcal{C}}} \log\left(h^*\left[Q_{\mathcal{C}}^{\mathcal{T}}\right](x_{\mathcal{C}})\right) \mathrm{SP}_{\mathcal{C}}(x_{\mathcal{C}})$$

where:

$$\mathrm{SP}_{\mathcal{C}}(x_{\mathcal{C}}) = \sum_{x\in\mathcal{X}_{X-\mathcal{C}}} \left(g[P](x_{\mathcal{C}}, x)\right)\left(h[Q](x_{\mathcal{C}}, x)\right) = \sum_{x\in\mathcal{X}_{X-\mathcal{C}}} \left[\prod_{\mathcal{C}\in\mathcal{C}_P} g\left[P_{\mathcal{C}}^{\mathcal{T}}\right](x)\right]\left[\prod_{\mathcal{C}\in\mathcal{C}_Q} h\left[Q_{\mathcal{C}}^{\mathcal{T}}\right](x)\right]$$

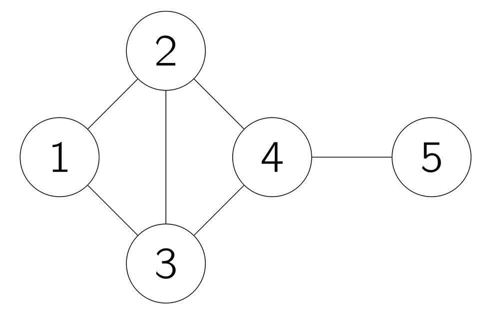
We can then obtain $\mathrm{SP}_{\mathcal{C}}(x_{\mathcal{C}})$ using belief propagation. For example:



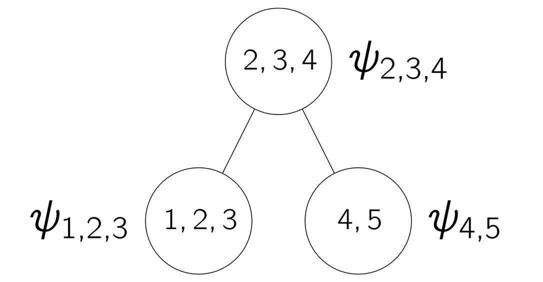


$$P = \frac{P_{1,2} \cdot P_{2,3,4} \cdot P_{4,5}}{P_2 \cdot P_4} = P_{2,3,4} \cdot P_{1|2} \cdot P_{5|4} \qquad Q = \frac{Q_{1,3} \cdot Q_{2,3,4} \cdot Q_{4,5}}{Q_2 \cdot Q_4} = Q_{2,3,4} \cdot Q_{1|3} \cdot Q_{5|4}$$

Computation graph $\mathcal{H}$



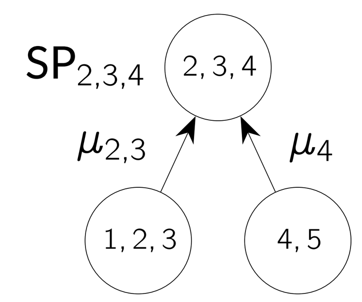
Clique tree of $\mathcal{H}$



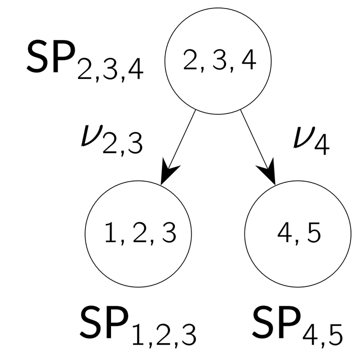
Initial factors

$$\psi_{2,3,4} = g[P_{2,3,4}] \cdot h[Q_{2,3,4}]$$
$$\psi_{1,2,3} = g[P_{1|2}] \cdot h[Q_{1|3}]$$
$$\psi_{4,5} = g[P_{5|4}] \cdot h[Q_{5|4}]$$

With $\mathcal{H}$ and the factors from $\mathrm{SP}_{\mathcal{C}}$ assigned to $\mathcal{C}(\mathcal{H})$, we can then proceed with carrying out belief propagation in order to obtain $\mathrm{SP}_{\mathcal{C}}$ for all $\mathcal{C} \in \mathcal{C}(\mathcal{H})$.
This involves propagating the factors assigned to the leaf cliques, up through the clique tree, to the root clique, then down back to the leaves.



$$\mu_{2,3}(x_{2,3}) = \sum_{x_1\in\mathcal{X}_1} \psi_{1,2,3}(x_1, x_{2,3})$$
$$\mu_4(x_4) = \sum_{x_5\in\mathcal{X}_5} \psi_{4,5}(x_5, x_4)$$
$$\mathrm{SP}_{2,3,4} = \psi_{2,3,4} \cdot \mu_{2,3} \cdot \mu_4$$



$$\nu_{2,3}(x_{2,3}) = \sum_{x_4\in\mathcal{X}_4} \mathrm{SP}_{2,3,4}(x_4, x_{2,3})$$
$$\nu_4(x_4) = \sum_{x_{2,3}\in\mathcal{X}_{2,3}} \mathrm{SP}_{2,3,4}(x_{2,3}, x_4)$$
$$\mathrm{SP}_{1,2,3} = \psi_{1,2,3} \cdot \nu_{2,3}/\mu_{2,3}$$
$$\mathrm{SP}_{4,5} = \psi_{4,5} \cdot \nu_4/\mu_4$$

Conclusion

In conclusion our approach:

- has a final complexity of:
$$D_{AB}^{(\alpha,\beta)}(P,Q) \in \begin{cases} \mathcal{O}(n^2 \cdot \omega 2^{\omega+1}) & \alpha,\beta = 0 \\ \mathcal{O}(n \cdot 2^{\omega(\mathcal{H})+1}) & \text{otherwise} \end{cases} \in \mathcal{O}\left(n^2\omega(\mathcal{H}) \cdot 2^{\omega(\mathcal{H})+1}\right)$$
- as fast as prev. approaches
- only requires belief propagation
- works on both Bayesian networks and Markov networks via conversion
- can compute a wider array of divergences other than the KL divergence

Runtime comparison with previous approach

Network	previous (secs)		ours (secs)	
	mean	sd	mean	sd
cancer	**0.0117**	0.0026	0.0132	0.0033
earthquake	0.0104	0.0025	**0.0075**	0.0009
survey	0.0140	0.0032	**0.0081**	0.0002
asia	0.0163	0.0001	**0.0137**	0.0007
sachs	0.0464	0.0106	**0.0151**	0.0001
child	0.0778	0.0101	**0.0402**	0.0013
insurance	0.3838	0.0051	**0.1590**	0.0029
water	**6.9326**	0.0329	7.6454	0.0637
mildew	**19.326**	0.1318	19.459	0.0852
alarm	0.3177	0.0099	**0.0875**	0.0018
hailfinder	0.8543	0.0243	**0.1672**	0.0052
hepar2	1.3058	0.0307	**0.2403**	0.0140
win95pts	1.0256	0.0289	**0.3538**	0.0049