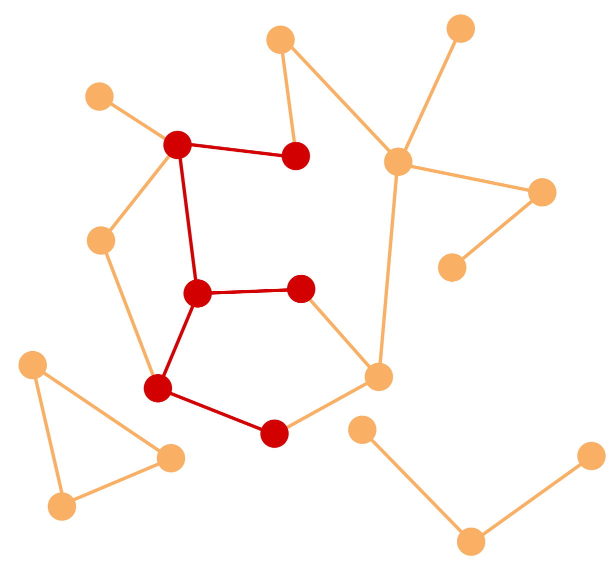
Understanding Domain-Size Generalization in Markov Logic Networks

Florian Chen, Felix Weitkämper and Sagar Malhotra

MOTIVATION

Relational data is becoming ever larger. Often only a subsample of the data is observed.

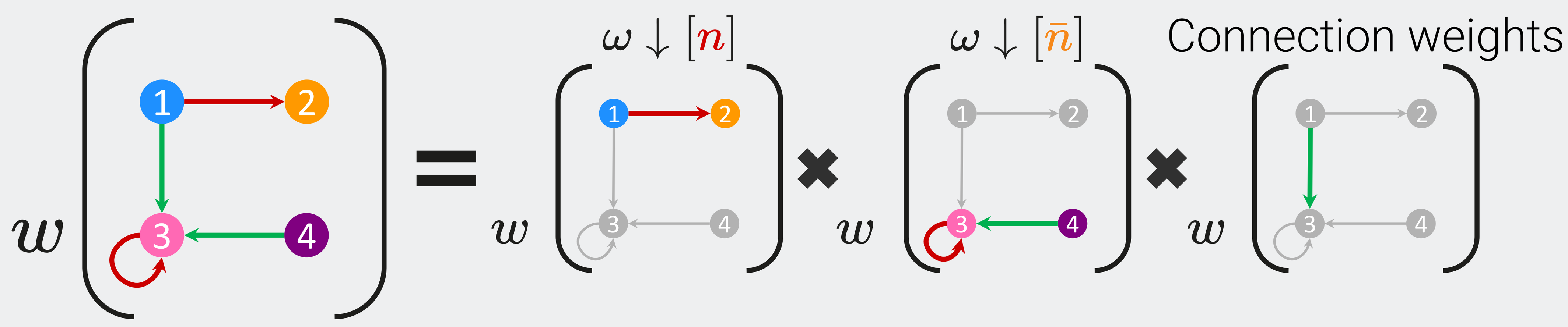


PROBLEM STATEMENT

Can we use a subsample to learn an accurate model for the larger structure?

WEIGHT DECOMPOSITION

We can decompose the weight of a structure as:



We can then use this decomposition to bound the weight of a structure:

$$w(\omega) \leq w(\omega \downarrow [n]) \times w(\omega \downarrow [\bar{n}]) \times \prod_{k \in [d]} (w_k^{max})^{\binom{n+m}{k} - \binom{n}{k} - \binom{m}{k}}$$

$$w(\omega) \geq w(\omega \downarrow [n]) \times w(\omega \downarrow [\bar{n}]) \times \prod_{k \in [d]} (w_k^{min})^{\binom{n+m}{k} - \binom{n}{k} - \binom{m}{k}}$$

MARKOV LOGIC

A Markov Logic Network (MLN) Φ is defined by weighted formulas $\{(\phi_i, a_i)\}_i$.

$$a_1 \quad \mathtt{Vaccine}(x) \Rightarrow \neg\mathtt{Covid}(x)$$
$$a_2 \quad \mathtt{Covid}(x) \wedge \mathtt{Contact}(x, y) \Rightarrow \mathtt{Covid}(y)$$

An MLN Φ induces a distribution over structures of size n:

$$P_\Phi^{(n)}(\omega) = \frac{1}{Z(n)} \exp\Big(\sum_{(\phi_i, a_i) \in \Phi} a_i N(\phi_i, \omega) \Big)$$

THEORETICAL CONTRIBUTION

$$M_{max} = \prod_{k \in [d]} (w_k^{max})^{\binom{n+m}{k} - \binom{n}{k} - \binom{m}{k}}$$
$$M_{min} = \prod_{k \in [d]} (w_k^{min})^{\binom{n+m}{k} - \binom{n}{k} - \binom{m}{k}}$$

Reducing parameter variance ...

... increases marginal likelihood

$$-\log P_\Phi^{(n+m)} \downarrow [n](\omega) \leq -\log P_\Phi^{(n)}(\omega) + \log \Delta$$

... decreases KL divergence

$$KL(P_\Phi^{(n+m)} \downarrow [n] || P_\Phi^{(n)}) \leq \log \Delta$$

$$\Delta = \frac{M_{max}}{M_{min}}$$

PARAMETER ESTIMATION

Maximum Likelihood (ML) estimate:

$$\hat{\mathbf{a}} = \operatorname*{argmax}_{\mathbf{a}} P_\Phi^{(n)}(\omega)$$

ML estimate for a subsample:

$$\hat{\mathbf{a}} = \operatorname*{argmax}_{\mathbf{a}} P_\Phi^{(n+m)} \downarrow [n](\omega)$$

However[1]:

$$\operatorname*{argmax}_{\mathbf{a}} P_\Phi^{(n)}(\omega) \neq \operatorname*{argmax}_{\mathbf{a}} P_\Phi^{(n+m)} \downarrow [n](\omega)$$

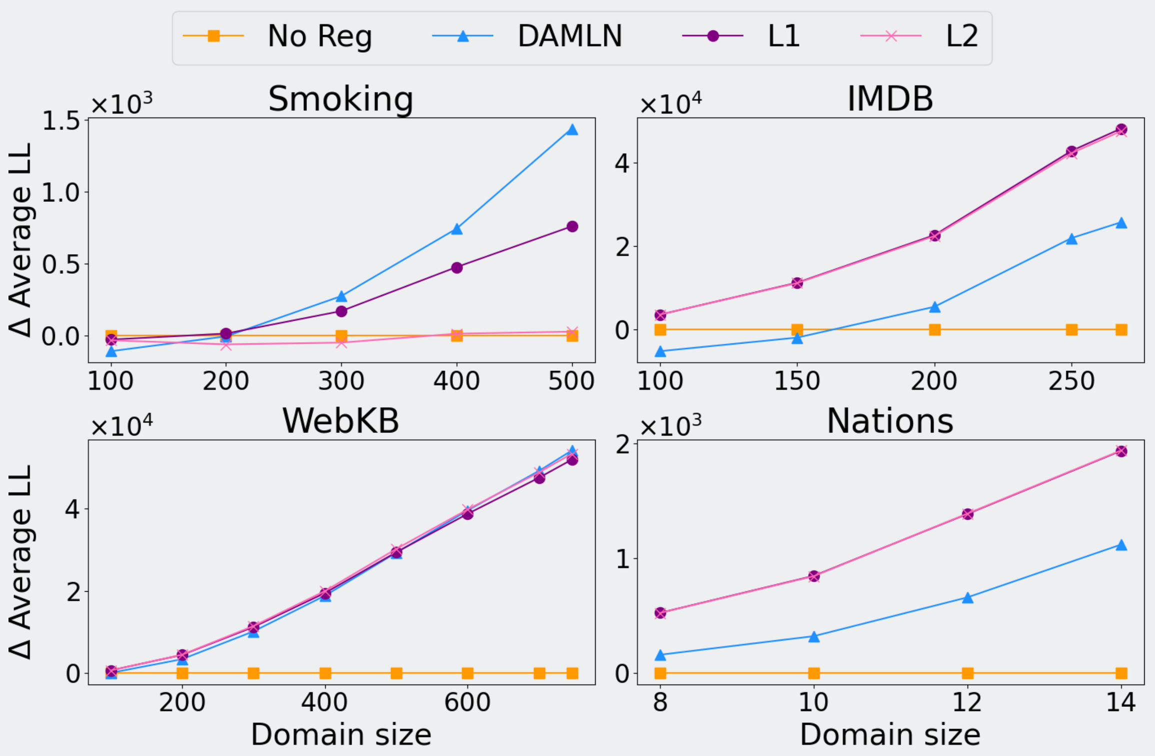
Can we analyze the relation between $P_\Phi^{(n)}$ and $P_\Phi^{(n+m)}$ to get better ML estimates for $P_\Phi^{(n+m)} \downarrow [n]$?

EXPERIMENTS AND RESULTS

We use **L1** and **L2** regularization, and Domain-Size Aware Markov Logic Networks[2] that downscale parameters:

$$P_\Phi^{(n)}(\omega) = \frac{1}{Z(n)} \exp\Big(\sum_{(\phi_i, a_i) \in \Phi} \frac{a_i}{s_i} N(\phi_i, \omega) \Big)$$

In the plots on the right, the average log-likelihood improvement across test domains of different sizes is displayed.



Regularization improves generalization.

Bibliography:
1. Shalizi, C.R., Rinaldo, A.: *Consistency under sampling of exponential random graph models* (2013)
2. Mittal, H., Bhardwaj, A., Gogate, V., Singla, P.: *Domain-Size Aware Markov Logic Networks* (2019)

Code Paper