Pattern-Guided Adaptive Prior for Structure Learning

Lyuzhou Chen, Yijia Sun, Yanze Gao, Xiangyu Wang*, Derui Lyu, Taiyu Ban,
Xin Wang, Xiren Zhou, Huanhuan Chen*

NeurIPS 2025

*corresponding authors

中国科学技术大学
University of Science and Technology of China

The Problem: "Correct but Imprecise" Priors are Harmful

Goal: Learning DAGs from observational data is hard, especially with limited data or noise.

Common Solution: Use prior knowledge (e.g., from domain experts).

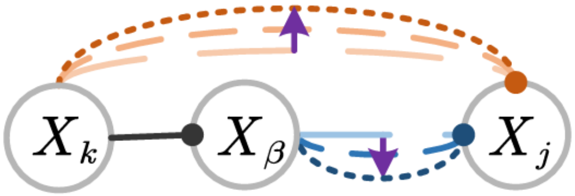
However, !!
Priors are often qualitative (e.g., "A influences B") , but models (like SEMs) require quantitative weights.

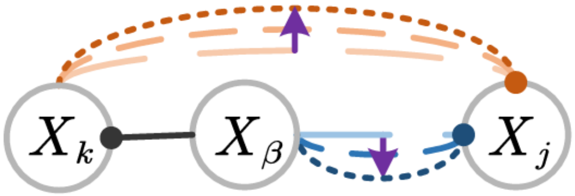
Forcing the model to fit these imprecise priors leads to **Edge Weight Deviation**.

Insight: Deviation Propagates and Creates Error Patterns

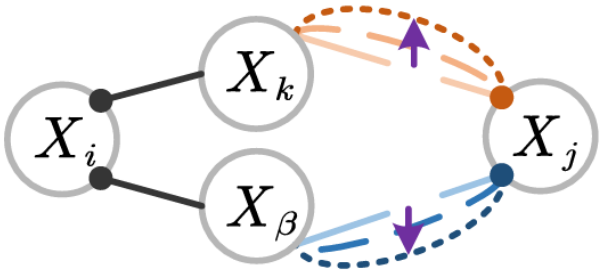
Key Point: This weight deviation is not isolated; it propagates and interferes with the learning of other edges.

We find This propagation creates two specific, observable structural error patterns.
Triangular Patterns :





and **Double Collision Patterns:**



These patterns are strong signals that a prior is causing harm.

Solution: The Pattern-Guided Adaptive Prior (PGAP) Framework

PGAP: A two-stage framework to leverage these error patterns.

Stage 1:
Learn an initial graph W_{stage1} using the priors

Stage 2: Adapt & Refine
1.Count the error patterns associated with each prior.
2.Adapt: If patterns are found, adaptively reduce the weight of that specific prior.
3.Refine: Re-learn the final graph with the adjusted weights.

Algorithm 1 PGAP with Soft Prior Constraints

```
Input: Data matrix X, set of prior knowledge Π, initial prior loss weights α_0 (vector with
  elements α_{0,π} for each π ∈ Π), pattern adjustment step size α.
Learn W_stage1 using a base structure learning algorithm, incorporating the prior loss term
  L_P(W; Π, α_0).
Initialize α_1 ← α_0.
for each prior π = (X_β → X_j) ∈ Π do
    n_patterns ← # of triangle and double collision patterns in W_stage1 involving π
    α_{1,π} ← max(0, α_{0,π} − α × n_patterns)
end for
Learn W_final using the base structure learning algorithm, incorporating the prior loss term
  L_P(W; Π, α_1).
Return W_final
```

Result: PGAP Improves Performance and Robustness

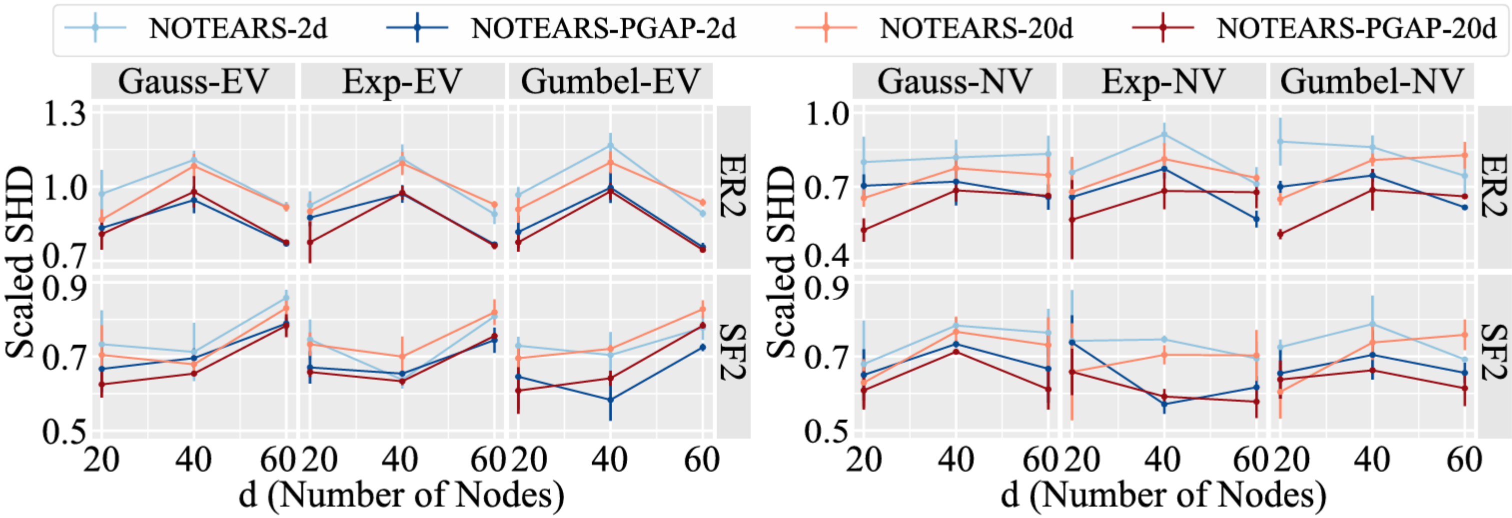


Figure 2: Performance of PGAP across Different Graph Settings (ER=2).

As shown in the figure, PGAP consistently achieves lower error (Scaled SHD) than the baseline. As shown in the charts, the baseline (lighter lines) can degrade as more priors are added. PGAP (darker lines) robustly improves, successfully mitigating the harm from imprecise priors.

Result: PGAP Enhances Existing SOTA Methods

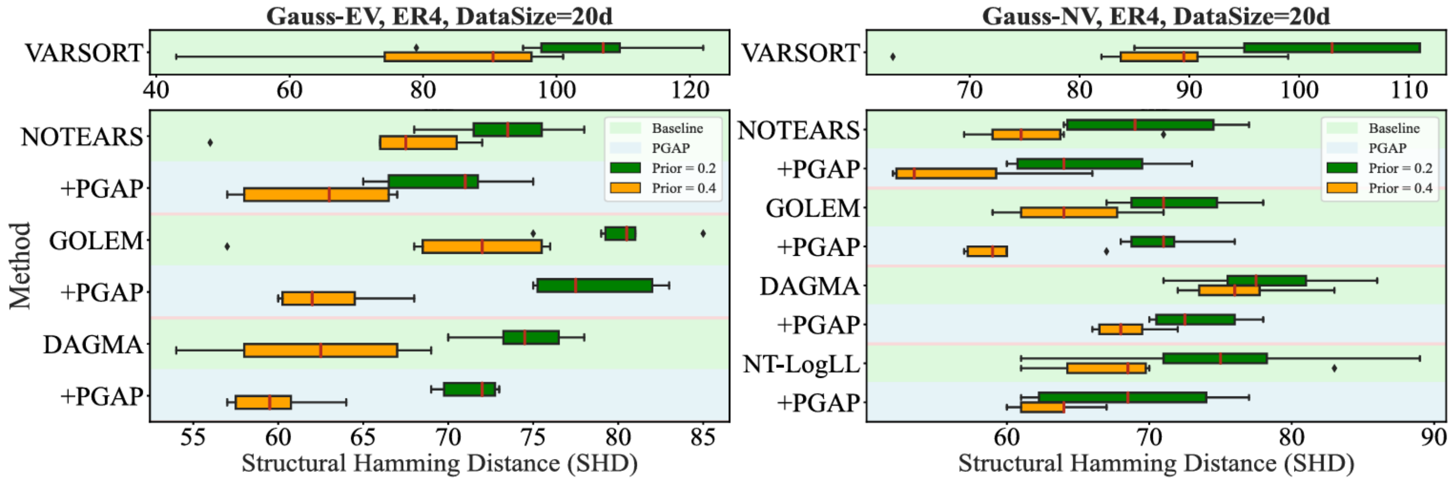


Figure 3: Improvements of PGAP on Mainstream Methods.

PGAP is not only a standalone algorithm, but also an adaptive "add-on" method.

As shown in the figure, PGAP enhances various existing methods, including NOTEARS, GOLEM, and DAGMA. The "+PGAP" versions consistently outperform their original counterparts (lower SHD).

PGAP Finds a Better Structure

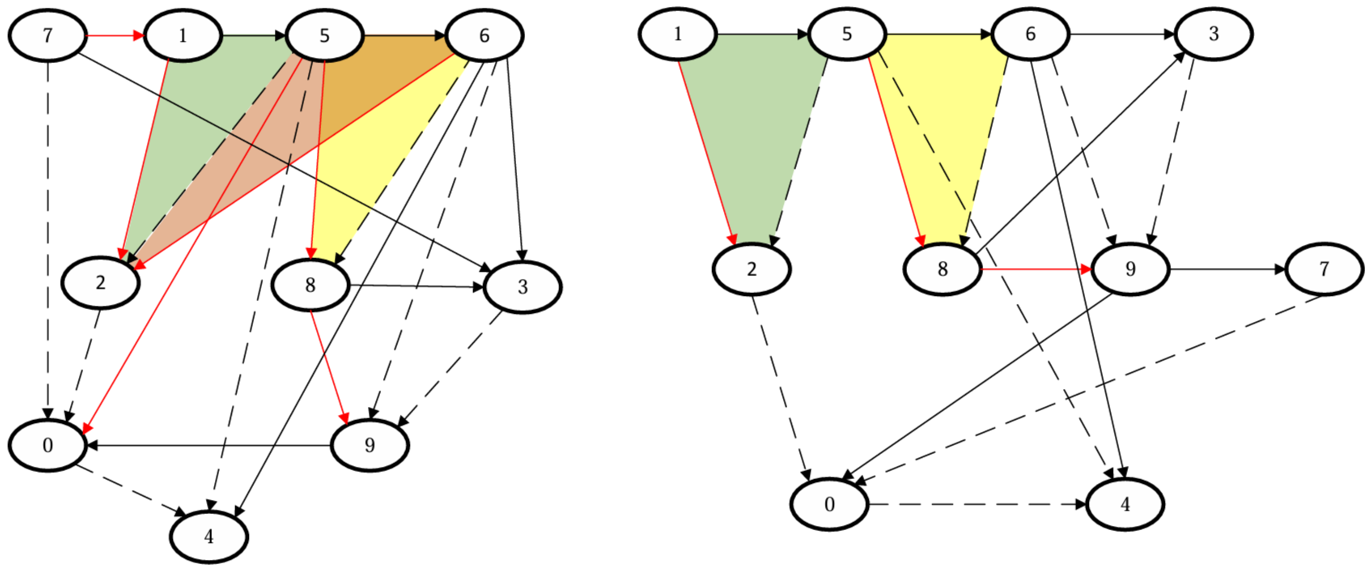


Figure 18: The structure learning results of NOTEARS-Softmax (left) and DAG-PGAP (right) for a DAG with 10 nodes, ER=2, and Gauss-NV noise. 40% priors are utilized.

Left (Baseline): The baseline method finds 6 incorrect "extra" edges (red lines). It also shows multiple error patterns (e.g., triangle 5-2-6).

Right (PGAP): Our PGAP-enhanced method finds only 3 incorrect edges.

PGAP successfully identified the pattern involving edge (6,2) and removed it.

Conclusion

Problem: Imprecise (qualitative) priors can cause edge weight deviation when used in quantitative models.

Insight: This deviation creates specific, observable error patterns (Triangular & Double Collision).

Solution: We proposed PGAP, a 2-stage framework that detects these patterns and adaptively reduces the influence of problematic priors.

Impact: PGAP provides a more robust and effective way to integrate prior knowledge in continuous optimization for structure learning.