Slide 1 – Title

My name is Safiye Sarac, and today I will present the key ideas and results of my master's thesis 'Selecting Candidate Parents for Bayesian Learning of DAGs'. This talk will explain why the candidate–parent problem matters, how we address it, and what empirical benefits our methods deliver.Thank you for joining."

Slide 2 – Contents

This presentation is structured in six parts: Introduction, Motivation, Contribution, Results, Limitations, and a brief Self-reflection.

Slide 3 – Introduction

We will introduce concept and termınoogy

Slide 4 – Directed Acyclic Graphs (DAGs)

A directed acyclic graph (DAG) is a directed graph that contains no directed cycles. Equivalently, its vertices can be arranged in a topological order in which every edge points from an earlier to a later vertex. This acyclicity property guarantees that the graph can represent causal or probabilistic flow without contradictions.

Slide 5 – Bayesian Network

A Bayesian network pairs a DAG with local conditional distributions. n a Bayesian network, the joint probability of all variables is obtained by multiplying the conditional probability of each variable given its parent variables in the network. This factorisation is what lets Bayesian networks scale to high-dimension.

Slide 6 – Bayesian Learning
 Bayesian learning uses Bayes' theorem to determine the conditional probability of a hypotheses given some evidence or observations. Our target is the posterior probability of a graph G after observing data D.
 'The posterior of G given D equals the marginal likelihood of the data under G times our prior belief in G, all divided by a normalising constant.'

Slide 7 – Motivation
"Learning structure is NP-hard: the number of DAGs grows super exponentially . Exhaustive algorithms stall around twenty-five variables."

Slide 8 – Why Select Candidate Parents?

Restricting each node to a small list of candidate parents shrinks the per-node search space by many orders of magnitude, making learning feasible.

SLIDE 9 Goals
That leads to two questions.

RQ 1 : Can we measure, quickly, how much posterior probability mass we lose when we select certian candidate parent set to K parents?
RQ 2 : Can we design heuristics that keep that loss tiny while scaling to hundreds of variables?"

Slide 10 – Challenge that motivate
One of the big motivation for selecting candidate parent sets is
 'The size of the DAG space explodes exponentially
Addressing these two gaps is crucial for scalable Bayesian learning.

. Worse, no principled metric had been available to quantify howe much posterior mass was lost by selecting a certain set of candidate parents ."
Slide 11 – Contributions (high-level)

I address those questions with four new heuristics and two probabilistic coverage metrics one fore when we have a fixed dataset therfore a fix post distributuion and one for we treat data and structure as a random variable."

Slide 12 – Existing Heuristics (Baselines)
"Before proposing new methods, I benchmarked eight previously studied heuristics Opt, Top-K, three Greedy variants, GES, PC, and Markov-Blanket. They trade speed for accuracy in different ways and serve as yard-sticks."

Slide 13 – Proposed Heuristics

Our contributions are Beam-BDeu, Post, Synergy, and Bootstrap-Stability all score-based, but each explores the parent-set lattice differently

Slide 14–15 – Approximate Posterior Coverage Fraction (APCF)
APCF metric is f
ormulated for a case where we fix a dataset so there is a fixed posterior distribution. This is relevant when
"APCF asks: 'What fraction of the posterior probability lies on graphs whose parent sets are already fully covered by our candidate lists?'

 In other words, it tells us how much mass survives the pruning.

Given a dag and a proposed candidate parent sets
 We say
for each node if proposed candidated parent set for the node is the superset of the parents of the node of the dag then that dag ıs covered . that we sample dags from the posterior given bdeu scores and compute the fraction of dags covered.

Slide 16–17 – Arc-Coverage (Random-data regime)

"When data and structure are both random, APCF becomes tiny and loses discrimination. I therefore defined Arc-Coverage—the recall of high posterior masds parent–child arcs on average over priors and data. It coincides with classical recall and stays informative .
"

We sample m number of ground truth dags from a fair structural prior and for each we sample data and parameters and we divide the number of edges that appear in ground truth dag that is proposed by heuristicsto the number of edges present in ground truth dag and we take the mean value.This gives us an expectation for our heuristic sucess.

Slide 18 – Results (Overview)

 Let's zoom into coverage resutls for previously mentioned 12 heuristics for two representative cases one for data is fixed one for we treat data as random variable ."
Slide 19 – Asia Network (visual)

"The eight-node Asia network is a classic benchmark with maximum in-degree 2.
It allows us to visualise the effect of K, the candidate-set size, on posterior coverage under controlled conditions."

Slide 20 – Asia Heat-map (APCF)

"This heat-map darker means higher APCF. Notice how the top-left remains pale for all methods; single-parent budgets are simply too small. But as soon as K crosses the ground truth maximum in-degree (here, 2), score-based methods surge, while CI-based ones lag two K-levels behind.At ten-thousand samples, Opt already covers fifty-eight percent of posterior mass at K = 3."

Slide 21 – Arc-Coverage (Random DAGs)

"Over 100 randomly drawn 10-node DAGs, the mean arc-coverage exceeds 0.99 at K = 7 for every score-based heuristic except Beam, which peaks slightly later. Constraint-based PC needs K = 9 to catch up, and Markov-Blanket never quite gets there."
Slide 22–23 – Limitations

"Our empirical study is limited to discrete Bayesian networks of up to 15 nodes in fixed-data tests and 10 nodes in simulations.
Continuous variables, larger graphs, and alternative structure priors remain open topics for future investigation."

Slide 24–25 – Self-Reflection

hroughout the project I applied four personal guidelines:

  No zero days—progress every day.

  Stay organised—version-control code and notes.

  Evaluate honestly—what went well, what did not.

  Enjoy the process—curiosity fuels perseverance

Slide 26 – References

"All formal proofs, algorithms, and additional plots are in the thesis manuscript and the Git repository linked there."
Slide 27 – Questions

"I'm happy to take your questions, comments, or suggestions."
Slide 28 – Thank You

"Thank you for listening."