outline for paper:

Start with sparsity boost idea:

BN structure learning is intractable for general distributions, but in practice we find that we can effectively learn the Bayesian network structure from real data distributions over relatively large numbers of variables (e.g. 100), This hints that there is some structure in real world distributions that may be aiding the structure learning process, and could possibly be leveraged in more targeted ways.

Techniques for BN structure learning can generally be categorized into either optimization approaches, where one defines some scoring function for each graph given the data, or constraint-based approaches, where the results of multiple conditional independence tests are pieced together to construct a graph structure.

Constraint-based approaches are sound in the sense that if one can accurately assess conditional independence, then one can recover the correct structure. Further, these techniques tend to be fast and scalable \cite{tsamardinos2006max}. However, they also tend to be brittle in the sense that one incorrect independence test can have global implications on the quality of the learned structure. Optimization approaches, given an appropriate scoring function, are also sound and avoid this brittleness, but face another problem, which is that the set of possible DAGs grows super-exponentially in the number of variables, V, i.e. O(2^V^2). Hence many resorted to heuristic search algorithms which could only guarantee local optimality