Supplementary Material to: Integer Programming for Causal Structure Learning in the Presence of Latent Variables

Heuristics for Separation at Fractional Solutions

The starting point of our heuristic is Karger's random contraction algorithm for finding near-optimal min-cuts in edge-weighted undirected graphs (with nonnegative weights). Given a weighted graph with n nodes and optimal min-cut value t and a positive integer $\alpha \geq 1$, Karger's algorithms runs in time bounded by a polynomial function of n^α and returns all cuts in the graph with weight $\leq \alpha t$. A weighted edge is chosen at random (with probability proportional to the weight of the edge), and the edge is contracted. When an edge ij is contracted where i and j are (pseudo-)nodes, let i' be a new pseudo-node representing $\{i, j\}$. Edges of the form ki or kj in the graph are removed and an edge ki' with weight $w_{ki'} = w_{ki} + w_{kj}$ is added, where w_{ki} is the weight of the edge ki in the graph before contraction and 0 if no such edge exists. This contraction procedure is repeated till there are 2α pseudo-nodes left, and the min-cut value in the resulting graph is returned. The central idea of the algorithm is that high weight edges are contracted resulting in the end-nodes of such edges being put in the same 'side' of the final cut.

We adapt the above idea. We first discuss how to find violated strengthened cluster inequalities. Consider a subset $S \subseteq V$ and a solution vector $\bar{z}$ of the LP relaxation. Let $\mu(S)$ equal the left-hand side of inequality where each z variable is set to the corresponding value in $\bar{z}$. If we find a subset $S \subset V$ such that $\mu(S) < 1$, then we have found a cluster inequality violated by the point $\bar{z}$. However, as there are exponentially many choices of the set S, it is not realistic to enumerate each S and compute $\mu(S)$. Instead, we initially only consider the sets $S = \{i\}$ consisting of individual nodes and note that $\mu(\{i\}) = 1$ for each node i because of equation $\sum_{C:i\in D_C} z_C = 1$. Let H_0 be the undirected weighted graph with the same set of nodes as G. We iteratively select and contract "high weight" edges and create pseudonodes (that consist of the union of nodes associated with the two pseudonodes incident to the edge), leading to a sequence of graphs $H_0, H_1, \ldots$, where each graph has one less pseudonode than the previous one. At the kth iteration we ensure that for each pseudonode $i \in H_k$, we have $\mu^k(\{i\})$ equal to the value of $\mu(S)$ where S is the set of nodes in H_0 that correspond to the pseudonode i of H_k.

Let the weight of an edge ij in H_0 be calculated as follows. Define

$$w_{ij} := \sum_{W:j\in W} \sum_{C\in\mathcal{C}:i\in D_C, W_{C,i}=W} \bar{z}_C + \sum_{W:i\in W} \sum_{C\in\mathcal{C}:j\in D_C, W_{C,j}=W} \bar{z}_C + \sum_{C\in\mathcal{C}:\{i,j\}\subseteq D_C, i\notin W_{C,j}, j\notin W_{C,i}} \bar{z}_C.$$

Note that the following relationship holds:

$$\mu(\{i, j\}) = \mu(\{i\}) + \mu(\{j\}) - w_{ij}. \tag{1}$$

Step 1: If we apply the random contraction step in Karger's algorithm to the weighted graph H_0 to obtain H_1, then with high probability we will contract an edge ij with a high value of w_{ij}. This step leads to an ij such that $\mu(\{i, j\})$ is approximately minimized (as $\mu(\{i\}) = \mu(\{j\}) = 1$ for all nodes i, j of H_0).

Step 2: We then create a pseudo-node $\{i, j\}$ in H_1 (labeled, say, by node i if $i < j$ and by j otherwise). Assuming the new psuedonode in H_1 has label i, We let $\mu^1(\{i\}) = \mu(\{i, j\})$ and $\mu^1(\{k\}) = \mu(\{k\})$ for all other nodes.
Step 3: We then recalculate w_{ij} values for edges in H_1 in such a fashion that for every pair of pseudonodes in H_1, the relationship in (1) holds. To do this, we first remove all c-component variables $\bar{z}_C$ where $i \in D_C$ and $j \in W_{C,i}$ or $j \in D_C$ and $i \in W_{C,j}$. Next we replace all occurrences of j by i in the remaining variables, and then recompute the weights w_{kl} for edges kl.

If we repeat Steps 1-3 for H_1 to obtain $H_2, H_3, \ldots$, then it is not hard to see that we always maintain the property in (1) with μ replaced by μ^k, and also the property that for any node i in H_k, the value $\mu^k(\{i\})$ is equal to $\mu(S)$ where S is the set associated with the pseudonode i. We stop whenever we find a pseudonode i in H_k (and associated S) such that $\mu^k(\{i\}) = \mu(S) < 1$. We repeat this algorithm multiple times while starting from different random seeds. Though this algorithm is not guaranteed to find a set S such that $\mu(S) < 1$, it works well in practice, and does not return spurious sets S.

To adapt the above algorithm to find violated strengthened bicluster inequalities, we proceed as follows. Consider a specific bidirected edge ij such that $\bar{I}(i \leftrightarrow j) > 0$ for a given fractional solution $\bar{z}$. We first contract ij in a special manner to obtain a graph H_0. Assume i' represents the resulting pseudonode: for any c-component C such that $i, j \in D_C$, we let $W_{C,i}$ and $W_{C,j}$ be replaced by a single parent set $W' = W_{C,i} \cup W_{C,j}$ of the new pseudonode i'. We also remove all c-component variables z_C such that $D_C \cap \{i, j\} = 1$. We subsequently define $\mu(\{k\})$ values for nodes k in H_0, edge weights w_{kl}, perform a random contraction step and repeat this process till we find a pseudonode i in H_k such that $\mu^k(\{i\}) < I(i \leftrightarrow j)$. We ensure that $\mu^k(\{i\})$ always represents the left-hand side of strengthened bicluster inequalities.

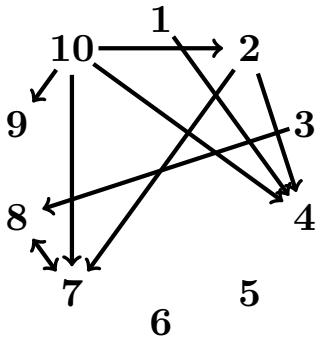
Performance of different methods when the number of latent variables increases

We present in the following table the precise numbers (means of SHD, precision and recall) of the results in Figure 6 of the main paper.

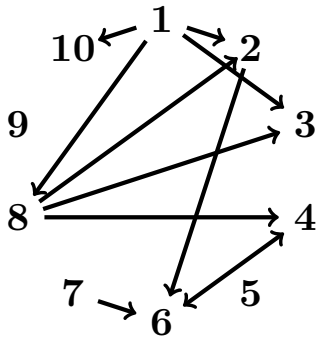
Table 1: Exact numbers for Figure 6

l	SHD				Precision (%)				Recall (%)			
	AGIP	M^{3}HC	FCI	cFCI	AGIP	M^{3}HC	FCI	cFCI	AGIP	M^{3}HC	FCI	cFCI
2	**12.6**	26.4	30.8	24.8	**80.3**	54.1	29.6	50.0	**78.7**	49.3	26.2	44.1
4	**22.4**	27.3	31.6	24.4	**63.0**	57.7	36.9	52.6	**63.1**	52.7	34.9	48.9
6	**28.8**	32.8	35.9	31.3	**57.8**	49.9	33.6	46.0	**53.1**	44.5	29.7	39.8

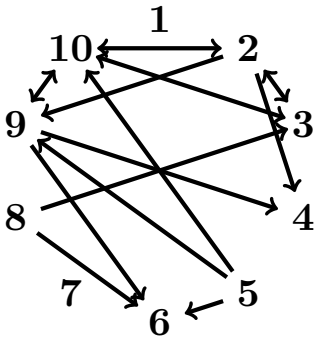
Ground Truth AGs for Experiments in Section 4.3 of the main paper



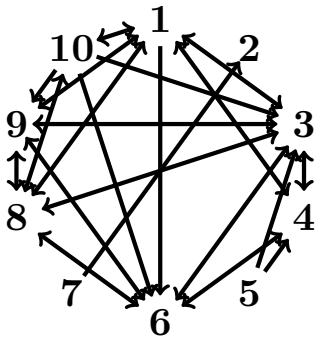
AG #1



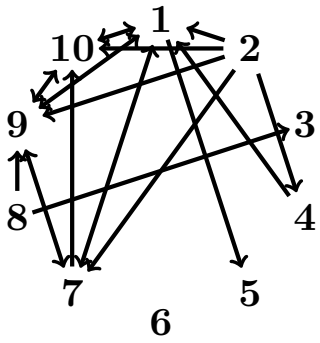
AG #2



AG #3



AG #4



AG #5