Supplementary Materials For: Faster algorithms for Markov equivalence

Zhongyi Hu
Department of Statistics
University of Oxford
zhongyi.hu@keble.ox.ac.uk

Robin Evans
Department of Statistics
University of Oxford
evans@stats.ox.ac.uk

A Proofs from Section 3

Lemma 3.5. If v, w are connected by a collider path π_1 in an ADMG $\mathcal{G}$ then they are connected by a collider path π_2 in $\mathcal{G}^m$ where π_2 uses a subset of the internal vertices of π_1. Also, if π_1 starts with $v \rightarrow$, so does π_2.

Proof. Any adjacent pair in $\mathcal{G}$ is also adjacent in $\mathcal{G}^m$ as any edge is a trivial collider path. So the path π_1 is still present in $\mathcal{G}^m$ however it may not be a collider path (if it is then we are done) and we aim to find a collider path π_2.

Suppose a is an internal vertex and is a noncollider in π_1 in $\mathcal{G}^m$ where $a \leftrightarrow b$ in $\mathcal{G}$ is changed to $a \rightarrow b$ in $\mathcal{G}^m$. This is because $a \in \mathrm{an}_{\mathcal{G}}(b)$. Consider the vertex c on the other side of a, suppose it is $c \leftrightarrow a$ in $\mathcal{G}$. Then $b \leftrightarrow a \leftrightarrow c$ is a collider path where $a \in \mathrm{an}_{\mathcal{G}}(\{b, c\})$ so b, c becomes adjacent in $\mathcal{G}^m$ and we can remove a from the path. If $c \rightarrow a$, i.e. c is one of end vertices, then in the projected graph we have $c \rightarrow a$. We can do this repeatedly until it terminates and the final path is a collider path in $\mathcal{G}^m$ that connects v, w. □

Lemma 3.7. Let v, w be two vertices then (i) $v \rightarrow w$ in $\mathcal{G}^m$ if and only if $v \in \mathrm{tail}_{\mathcal{G}}(w)$ and (ii) $v \leftrightarrow w$ in $\mathcal{G}^m$ if and only if $\{v, w\} \in \mathcal{H}(\mathcal{G})$.

Proof. For (i), if $v \rightarrow w$ in $\mathcal{G}^m$ then $v \in \mathrm{an}_{\mathcal{G}}(w)$ and in $\mathcal{G}$ there is an inducing path between v and w (a collider path). If $v \rightarrow w$ in $\mathcal{G}$ then we are done. Otherwise any intermediate vertex on the path is in $\mathrm{an}_{\mathcal{G}}(\{v, w\}) = \mathrm{an}_{\mathcal{G}}(w)$ hence $v \mathrel{?\!\!\rightarrow} \ldots \leftrightarrow w$. Therefore $v \in \mathrm{tail}_{\mathcal{G}}(w)$. Conversely, $v \in \mathrm{tail}_{\mathcal{G}}(w)$ implies that $v \in \mathrm{an}_{\mathcal{G}}(w)$ and there is a collider path between v and w with any intermediate vertex in $\mathrm{an}_{\mathcal{G}}(w)$ hence the path is an inducing path and $v \rightarrow w$ in $\mathcal{G}^m$.

For (ii), if $v \leftrightarrow w$ in $\mathcal{G}^m$ then there is an inducing path between v and w (a collider path) in $\mathcal{G}$ and v, w are not ancestors to each other. Also any intermediate vertex on the path is in $\mathrm{an}_{\mathcal{G}}(\{v, w\})$ which suggests that the path is a bidirected path. Therefore, $\{v, w\}$ forms a head. On the other hand, if $\{v, w\}$ is a head in $\mathcal{G}$ then they are not ancestors to each other and there is a bidirected path between them with any intermediate vertex in $\mathrm{an}_{\mathcal{G}}(\{v, w\})$ so this path is an inducing path and $v \leftrightarrow w$ in $\mathcal{G}^m$. □

Theorem 3.8. For two ADMGs $\mathcal{G}_1$ and $\mathcal{G}_2$, they are ordinary Markov equivalent if and only $\mathcal{S}(\mathcal{G}_1) = \mathcal{S}(\mathcal{G}_2)$.

Proof. This follows from Proposition 3.7 and Theorem 3.2. □

Corollary 3.8.1. Two ADMGs $\mathcal{G}_1$ and $\mathcal{G}_2$ are ordinary Markov equivalent if and only $\mathcal{S}_3(\mathcal{G}_1) = \mathcal{S}_3(\mathcal{G}_2)$, and this occurs if and only if $\tilde{\mathcal{S}}_3(\mathcal{G}_1) = \tilde{\mathcal{S}}_3(\mathcal{G}_2)$.

Proof. By Proposition 3.8, $\mathcal{S}_3$ are preserved in $(\mathcal{G}_1)^m$ and $(\mathcal{G}_2)^m$, and with the new definition of adjacencies, the outputs of $\tilde{\mathcal{S}}_3$ are also preserved. Hence the statement follows from Corollary 3.2.1. □

B Extension to Summary Graphs and MAGs with undirected edges

MAGs defined in Richardson and Spirtes (2002) contain undirected edges which necessitate additional conditions of ancestrality. In addition to the previous condition ($\mathrm{sib}_{\mathcal{G}}(v) \cap \mathrm{an}_{\mathcal{G}}(v) = \emptyset$ and this is referred as condition 1 of ancestrality), one also requires that if an undirected edge is present between two vertices v and w then there is no arrow into v or w. We refer to this as condition 2 of ancestrality.

Definition B.1. A graph $\mathcal{G}$ is *ancestral* if: (1) for every $v \in \mathcal{V}$, $\mathrm{sib}_{\mathcal{G}}(v) \cap \mathrm{an}_{\mathcal{G}}(v) = \emptyset$; (2) if there is an undirected edge $x - y$ then x, y have no parents and no siblings.

A direct consequence of this definition is that vertices with undirected edges are 'at the top' of the graph $\mathcal{G}$. For an acyclic graph $\mathcal{G}$ with three types of edges and only

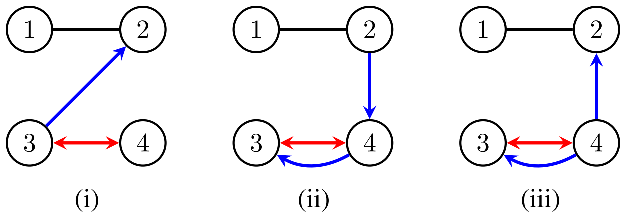

Figure 1: (i) A graph that satisfies only condition 1 of ancestrality. (ii) A graph that satisfies only condition 2 of ancestrality. (iii) A graph that does not satisfy either condition 1 or 2 of ancestrality.

satisfying condition 2 of ancestrality, it can be seen as an ADMG with an undirected component among vertices without parents or siblings and therefore the component is "at the top" of the graph.

Summary graphs defined in Wermuth (2011) are actually the same as ADMGs with undirected components at the top. Graphically, one just needs to change the dashed lines to bidirected edges and they encode the same conditional independence. For simplicity, we will refer to this type of graphs as summary graphs. Among the three graphs in Figure 1, (ii) is the only summary graph.

Definition B.2. For a summary graph $\mathcal{G}$, let $U = \{v \in \mathcal{V} : v - w \text{ for some } w \in \mathcal{V}\}$ and $D = \mathcal{V} \setminus U$. Define $\mathcal{G}^u = \mathcal{G}_U$ and $\mathcal{G}^d = \mathcal{G}_D$.

It is showed by Richardson and Spirtes (2002) that we can always split a summary graph into two disjoint subgraphs. One is an undirected subgraph $\mathcal{G}^u$ and another one is a subgraph with only directed and bidirected edges $\mathcal{G}^d$. Note that heads and barren sets are only defined in $\mathcal{G}^d$, and tails may include vertices in both $\mathcal{G}^u$ and $\mathcal{G}^d$.

For example, Figure 1(ii) can be split as $\mathcal{G}^u = 1 - 2$ and $\mathcal{G}^d = 3 \leftarrow 4, 3 \leftrightarrow 4$. Its heads are $\{3\}$ and $\{4\}$ and the corresponding tails are $\{2, 4\}$ and $\{2\}$.

A vertex a is said to be *anterior* to b if there is a path π on which every edge is either undirected or directed towards b, or if $a = b$. We denote the collection of all vertices anterior to b by $\text{ant}_{\mathcal{G}}(b)$.

An *undirected graph* (UG) is a graph with only undirected edges. A *clique* in an UG is defined as a complete subset of vertices, that is: every pair of vertices is connected by an undirected edge.

For summary graphs, including MAGs, a *clique* is defined in the same manner for vertices in $\mathcal{G}^u$, with completeness referring only to adjacencies by undirected edges.

Remark. We extend the definition of parametrizing set by adding all the cliques to the set.

B.1 Extension to MAGs With Undirected Edges

We only need to add a few line of argument to extend previous propositions and theorems.

For $\Rightarrow$ of Proposition 3.3: if $W \in \mathcal{S}(\mathcal{G})$, then either W is a clique or there is a nonempty subset $W' \subseteq W$ such that W' is a head and $W \subseteq W' \cup \text{tail}(W')$. The latter case is proved in the main paper. For the former case, it clearly implies that we can not m-separate any two vertices in W, given the remaining vertices in W.

For $\Leftarrow$ of Proposition 3.3: For W that does not lie entirely in $\mathcal{G}^u$ we can define $W' = \text{barren}(W)$. For W lying in $\mathcal{G}^u$, if we cannot m-separate any two vertices in W then clearly W is a clique and $W \in \mathcal{S}(\mathcal{G})$.

Proposition 3.4 does not change if we add undirected edges in MAGs, thus Theorem 3.2 and Corollary 3.2.1 hold for MAGs with undirected edges.

B.2 Extension to Summary Graphs

The projection described in Section 3.3 can be extend to summary graphs with latent variables L as stated in Richardson and Spirtes (2002). The modified projection is: (i) every pair of vertices $a, b \in \mathcal{V}$ in $\mathcal{G}$ that are connected by an *inducing path* becomes adjacent in $\mathcal{G}^m$; (ii) an edge connecting a, b in $\mathcal{G}^m$ is oriented as follows: if $a \in \text{ant}_{\mathcal{G}}(b)$ then $a \rightarrow b$; if $b \in \text{ant}_{\mathcal{G}}(a)$ then $b \rightarrow a$; if neither is the case, then $a \leftrightarrow b$; if they are both anterior to one another then the edge is undirected. An *inducing path* between a, b is a path such that every collider in the path is in $\text{an}(\{a, b\})$, and every noncollider is in L. Again, we only consider projections with no latent variable, so an inducing path is just a collider path with every collider in $\text{an}(\{a, b\})$. And the projection still preserves ancestral relations from the original graph. We first show that undirected edges are preserved through projections.

Lemma B.1. *If $\mathcal{G}$ is a summary graph and $\mathcal{G}^m$ is its corresponding projected MAG, then $\mathcal{G}^u = (\mathcal{G}^m)^u$ and $(\mathcal{G}^d)^m = (\mathcal{G}^m)^d$.*

Proof. For the first statement, we can prove it by showing that undirected edges are the same. First of all, notice that all undirected edges in $\mathcal{G}$ is preserved in $\mathcal{G}^m$. Secondly, no additional undirected edges can be added. If a and b are both in $\mathcal{G}^u$ then if they are not adjacent before, they are still nonadjacent since there is no inducing path between them (they are already at the top of the graph). If a and b are both in $\mathcal{G}^d$ then they cannot be anterior to each other, this would violate condition (ii) of ancestrality or the fact that $\mathcal{G}$ is acyclic. If $a \in \mathcal{G}^u$ and $b \in \mathcal{G}^d$ then obviously b cannot be anterior to a.

For the second statement, note the two subgraphs have the same vertices due to the first statement. For vertices in $\mathcal{G}^d$, ancestral relations are the same in $\mathcal{G}$ as there is no directed path passing $\mathcal{G}^u$. Also when we consider inducing paths, any such path would not contain any vertex in $\mathcal{G}^u$. □

We now show that Proposition 3.6 also holds for summary graphs, i.e. heads and tails are preserved through projection.

Proof. So we have proved that for ADMGs, heads and tails are preserved through the projection. Now heads are only defined in $\mathcal{G}^d$ and $(\mathcal{G}^m)^d$, thus by Lemma $B.1$, for a summary graph, heads are preserved in $\mathcal{G}^m$. Also for tails that are in $\mathcal{G}^d$, they are preserved. It remains to show that the result holds when tails are in $\mathcal{G}^u$. For a head H, let $w \in \mathcal{G}^u$. If $w \in \mathrm{tail}_{\mathcal{G}}(H)$ then we know there is a path $\pi : w \to w_1 \leftrightarrow \cdots \leftrightarrow h$, for $h \in H$ with intermediate vertices in $\mathrm{an}(H)$. Although $w \notin \mathcal{G}^d$, with the same argument in Lemma 3.5, this path is preserved as a collider path in $\mathrm{an}(H)$ in $\mathcal{G}^m$ with $\leftrightarrow h$ (h is in a head) hence $w \in \mathrm{tail}_{\mathcal{G}^m}(H)$. Suppose now $w \in \mathrm{tail}_{\mathcal{G}^m}(H)$, so there is a path $\pi : w \to w_1 \leftrightarrow \cdots \leftrightarrow h$ with intermediate vertices in $\mathrm{an}(H)$, we know every bidirected edge corresponds to a bidirected path in $\mathrm{an}(H)$ in $\mathcal{G}$, and the first directed edge correspond to a path $\pi' : w \to w_1 \leftrightarrow \cdots \leftrightarrow w_2$ in $\mathcal{G}$ with intermediate vertices in $\mathrm{an}(w_2) \subseteq \mathrm{an}(H)$, thus $w \in \mathrm{tail}_{\mathcal{G}}(H)$. □

Since Proposition 3.6 holds for summary graphs, if we change the definition of *adjacencies* in summary graphs in the same manner as ADMGs by referring to m-separations, Theorem 3.8 and Corollary 3.8.1 also hold for summary graphs.

B.3 Extension for Algorithms

For Algorithm 1, we only add a line at the end of the algorithm (after line 17) to obtain the connected pairs in $\mathcal{G}^u$ (referred as line 18 in the next section). This costs $O(e)$ and hence does not contribute to the overall complexity.

For Algorithm 2, as showed by Lemma B.1, undirected edges are preserved, it is sufficient to add a line at the end of the algorithm (after line 9) to keep all the undirected edges. This costs $O(e)$ and hence does not contribute to the overall complexity.

C Proof that Algorithm 1 outputs $\tilde{\mathcal{S}}_3$

Let $A_1(\mathcal{G})$ be the output of Algorithm 1 and $A_1'(\mathcal{G})$ be the output of Algorithm 1 without checking adjacencies in lines 6, 11 and 14. We also define the following sets for a MAG $\mathcal{G}$:

$$
\begin{aligned}
H_1(\mathcal{G}) &= \{\{v,w,z\} : v \in \mathcal{V} \text{ and } w,z \in \mathrm{pa}_{\mathcal{G}}(v)\}\\
H_2(\mathcal{G}) &= \{\{v,w,z\} : v \leftrightarrow w, z \in \mathrm{tail}(\{v,w\})\}\\
H_3^a(\mathcal{G}) &= \text{all heads of size 3 with some adjacencies}\\
H_3^n(\mathcal{G}) &= \text{all heads of size 3 with no adjacencies}\\
H_3(\mathcal{G}) &= \text{all heads of size 3} = H_3^a(\mathcal{G}) \cup H_3^n(\mathcal{G})\\
\hat{\mathcal{S}}_3(\mathcal{G}) &= \{S \in \mathcal{S}_3(\mathcal{G}) : \text{there are some adjacencies in } S\}\\
U_3(\mathcal{G}) &= \{S \subseteq \mathcal{V}(\mathcal{G}^u) : |S| = 3 \text{ and } S \text{ is complete}\}.
\end{aligned}
$$

Thus by definition $\tilde{\mathcal{S}}_3(\mathcal{G}) \subseteq \hat{\mathcal{S}}_3(\mathcal{G}) \subseteq \mathcal{S}_3(\mathcal{G})$ and $\mathcal{S}_2(\mathcal{G})$, $H_1(\mathcal{G})$, $H_2(\mathcal{G})$, $H_3^a(\mathcal{G})$, $H_3^n(\mathcal{G})$, $U_3(\mathcal{G})$ are disjoint.

Lemma C.1. *In a MAG $\mathcal{G}$, for any single vertex a, $\mathrm{tail}(a) = \mathrm{pa}_{\mathcal{G}}(a)$, and $\{v,w\}$ is a head if and only if $v \leftrightarrow w$.*

Proof. If $a \subset \mathrm{dis}_{\mathrm{an}(a)}(a)$ then there is a vertex b such that $b \leftrightarrow a$ and $b \in \mathrm{an}_{\mathcal{G}}(a)$, which contradicts ancestrality. Hence $\mathrm{tail}(a) = \mathrm{pa}_{\mathcal{G}}(a)$.

If $v \leftrightarrow w$ then v, w have no ancestral relation so by definition, it is a head. Suppose $\{v,w\}$ is a head, so $\{v,w\} \in \mathcal{S}(\mathcal{G})$ then they must be adjacent by Proposition 3.4 and the adjacency can not be undirected or directed, thus $v \leftrightarrow w$. □

Thus $H_1(\mathcal{G})$ and $H_2(\mathcal{G})$ are precisely the sets in $\mathcal{S}_3(\mathcal{G})$ that arise from heads of size one and two, respectively.

Lemma C.2. *For a MAG $\mathcal{G}$, we have*

$$
\begin{aligned}
\mathcal{S}_3(\mathcal{G}) &= \mathcal{S}_2(\mathcal{G}) \cup H_1(\mathcal{G}) \cup H_2(\mathcal{G}) \cup H_3(\mathcal{G}) \cup U_3(\mathcal{G})\\
\hat{\mathcal{S}}_3(\mathcal{G}) &= \mathcal{S}_2(\mathcal{G}) \cup H_1(\mathcal{G}) \cup H_2(\mathcal{G}) \cup H_3^a(\mathcal{G}) \cup U_3(\mathcal{G}).
\end{aligned}
$$

Proof. Consider the first equality, for $S = \{v,w\} \in \mathcal{S}_3(\mathcal{G})$, by Proposition 3.4, v, w are adjacent in $\mathcal{G}$ so $S \in \mathcal{S}_2$; For $S \in \mathcal{S}_3(\mathcal{G})$ and $|S| = 3$, it is a clique in $\mathcal{G}^u$ or it origins from heads of size either 1 or 2 or 3. Thus by Lemma 4.1 and Lemma 4.1, $S \in H_1(\mathcal{G}) \cup H_2(\mathcal{G}) \cup H_3(\mathcal{G}) \cup U_3(\mathcal{G})$; For S in the right hand side, it is in $\mathcal{S}_3(\mathcal{G})$ by definition.

For the second equality, by definition $\hat{\mathcal{S}}_3(\mathcal{G})$ excludes all $S \in \mathcal{S}_3(\mathcal{G})$ that have no adjacencies, but note that all $S \in \mathcal{S}_2(\mathcal{G}) \cup H_1(\mathcal{G}) \cup H_2(\mathcal{G}) \cup U_3(\mathcal{G})$ have some adjacencies. And by definition $H_3^a(\mathcal{G})$ extract all heads of size 3 with some adjacencies. □

Lemma C.3. *For a MAG $\mathcal{G}$, $A_1'(\mathcal{G}) \cup U_3(\mathcal{G}) = \hat{\mathcal{S}}_3(\mathcal{G})$.*

Proof. $S \in A_1'(\mathcal{G})$ obtained at line 5, 7, 9, 12, 17 and 18, correspond to sets in $\mathcal{S}_2(\mathcal{G})$, $H_1(\mathcal{G})$, $\mathcal{S}_2(\mathcal{G})$, $H_2(\mathcal{G})$, $H_3^a(\mathcal{G})$ and $\mathcal{S}_2(\mathcal{G})$, respectively. So by Lemma C.2, $A_1'(\mathcal{G}) \cup U_3(\mathcal{G}) \subseteq \hat{\mathcal{S}}_3(\mathcal{G})$ Conversely, all sets in $\hat{\mathcal{S}}_3(\mathcal{G}) \setminus U_3(\mathcal{G})$ can be obtained at corresponding lines. □

Proposition 4.1. For a MAG $\mathcal{G}$, $A_1(\mathcal{G}) = \tilde{\mathcal{S}}_3(\mathcal{G})$.

Proof. Compared to $\hat{\mathcal{S}}_3(\mathcal{G})$, $\tilde{\mathcal{S}}_3(\mathcal{G})$ excludes all sets of size 3 that have 3 adjacencies. If the set is clique in $\mathcal{G}^u$ except for edges, it is not added in Algorithm 1. Otherwise note that when sets of size 3 are obtained, lines 6, 11 and 14 check their adjacencies. $\square$

Notice that Algorithm 1 naturally identifies $\hat{\mathcal{S}}_3(\mathcal{G}) \setminus U_3(\mathcal{G})$, but to obtain the full $\hat{\mathcal{S}}_3(\mathcal{G})$ one also needs to identify all triangles in the undirected component; $\tilde{\mathcal{S}}_3(\mathcal{G})$ excludes this set.

Proposition 4.2. Let A_i be the number of ancestors of the vertex i. Then

$$\mathbb{E}A_i = \left(1 + \frac{r}{n}\right)^{i-1}.$$

In particular,

$$\mathbb{E}A_n = \left(1 + \frac{r}{n}\right)^{n-1} \longrightarrow e^r.$$

Proof. We proceed by induction. The result is trivially true for $A_2 = 1 + \frac{r}{n}$. Suppose the result holds for A_j. Then

$$\begin{aligned}\mathbb{E}A_{j+1} &= 1 + \sum_{i=1}^{j} \mathbb{E}\mathbb{1}_{\{i \to j+1\}} A_i \\ &= 1 + \frac{r}{n} \sum_{i=1}^{j} \left(1 + \frac{r}{n}\right)^{i-1},\end{aligned}$$

using independence of the edge and A_i and the induction hypothesis. Hence

$$\begin{aligned}\mathbb{E}A_{j+1} &= 1 + \sum_{i=1}^{j} \sum_{k=0}^{i-1} \binom{i-1}{k} \left(\frac{r}{n}\right)^{k+1} \\ &= 1 + \sum_{k=0}^{j-1} \left(\frac{r}{n}\right)^{k+1} \sum_{i=k+1}^{j-1} \binom{i-1}{k} \\ &= 1 + \sum_{k=0}^{j-1} \left(\frac{r}{n}\right)^{k+1} \binom{j}{k+1} \\ &= 1 + \sum_{k=1}^{j} \left(\frac{r}{n}\right)^{k} \binom{j}{k}.\end{aligned}$$

by a standard result about binomial coefficients. This gives the result. $\square$

References

T. S. Richardson and P. Spirtes. Ancestral graph Markov models. *Annals of Statistics*, 30(4):962–1030, 08 2002.

N. Wermuth. Probability distributions with summary graph structure. *Bernoulli*, 17(3):845–879, 08 2011.