SUPPLEMENTARY MATERIAL
IDENTIFYING THE TYPES OF CAUSES USING OBSERVATIONAL DATA

A PREPRINT

Author
Affiliation
Address
`email`

Coauthor
Affiliation
Address
`email`

Coauthor
Affiliation
Address
`email`

February 20, 2019

A Algorithms

A.1 Algorithm for Finding Chordless Path

Algorithm 4, which gives a method to find all chordless paths with X as their one endpoint in a given chordal graph, is the basis of Algorithm 1. Algorithm 4 is a variant of Breadth-First-Search. It starts from X, by adding $\{(X, \emptyset, 0)\}$ to a waiting queue S. Each triple in S consists of T, P, and n. The first element, T, is the vertex we will visit, which stands for an endpoint of a chordless path starting from X denoted by $\pi(X, T)$, the second element P is the vertex adjacent to T on $\pi(X, T)$, and the third element n is the unique ID of the subpath $\pi(X, P)$. During the k-th loop, we take the first element (T, P, n) out of S and add (T, P, n, k) to the front of F, which stores all the desired paths that we have already found. Each quadruple in F contains 4 elements, namely T, P, n and k. The first three elements have the meanings as in S, while the forth element k means the current number of loops, which is also the ID of the chordless path $\pi(X, T)$. Next, we extend the current path $\pi(X, T)$ to a longer chordless path by appending a vertex adjacent to T but not adjacent to P. More precisely, let α be a such vertex, the chordless path $\pi(X, T)$ and the edge $T - \alpha$ generate a new chordless path $\pi(X, \alpha)$. We label $\pi(X, T)$ by number k and add (α, T, k) to the end of S. Of course, such a vertex α may not exist and if so, no triple will be added to S.

Basically, The output F is a linearization of a tree. Given the output queue F of Algorithm 4, each element in F except for $(X, \emptyset, 0, 1)$ represents a chordless path that starts with X. More specifically, let $(T, P, n, k) \in F$ and $T \neq X$, (T, P, n, k) represents a chordless path starting with X and ending with T. To recover this path, we first add T into a queue, and find the quadruple after (T, P, n, k) in F whose forth element is n. Clearly, the first element of this quadruple is P. Again, add P into the queue. Repeat the above procedures until we reach $(X, \emptyset, 0, 1)$ and add X into the queue. Finally, reverse the queue to get a chordless path starting with X and ending with T.

It can be shown that,

Lemma 2. *The output of Algorithm 4 consists of all the chordless paths that start with X in a given chordal graph $\mathcal{C}$.*

Algorithm 4 Find every chordless path starting with X in a given chordal graph $\mathcal{C}$

Require: A chordal graph $\mathcal{C}$, X
Ensure: A sequence F that store all the chordless paths that start with X

initializing queues $S = \{(X, \emptyset, 0)\}$ and $F = \emptyset$,
k=0,
while S is not empty **do**
    k=k+1,
    take the first element (T, P, n) out of S and add (T, P, n, k) to the front of F,
    if $P \neq \emptyset$ **then**
        for $\alpha \in adj(T)$ such that $\alpha \neq P$ **do**
            if $\alpha \notin adj(P)$ **then**
                add (α, T, k) to the end of S,
            end if
        end for
    else
        for $\alpha \in adj(T)$, **do**
            add (α, T, k) to the end of S,
        end for
    end if
end while
return F.

B Detailed Proofs

More concepts and lemmas are needed before we present the proofs of Lemmas, Theorems and Corollaries from our main text.

Let $\mathcal{C} = (V, E)$ be an undirected graph with vertex set V and edge set E. Given $S \subseteq E$, the subgraph $\mathcal{C}(S)$ of $\mathcal{C}$ induced by S is an undirected graph with vertex set S and edge set $E(S) = \{(u, v) \in E | u, v \in S\}$. An induced subgraph $\mathcal{C}(S)$ is called complete if there is an edge between every pair of distinct nodes in S. A clique is a set of vertices such that the induced subgraph is complete. Moreover, a clique is maximal if it isn't contained in any other cliques.

Let $\pi = (v_0, v_1, ..., v_k)$ denote a simple path with length equals to k. If $k \geq 2$, we say three consecutive vertices v_i, v_{i+1} and v_{i+2} form a triangle on π if v_i is adjacent to v_{i+2}. π is called triangle-free if it doesn't contain any triangle. It can be shown that,

Lemma 3. *In any chordal graph, a path is chordless if and only if it's triangle-free.*

Proof. Let $\pi = (v_0, v_1, ..., v_k)$ denote a simple path with length $k \geq 2$. If π is chordless, then it's obviously triangle-free. Now suppose π is not chordless, then we can pick up a chord $v_i - v_j$ such that the subpath $\pi(v_i, v_j)$ has no chord except for $v_i - v_j$. If $j = i + 2$, then v_i, v_{i+1} and v_j form a triangle. If $j > i + 2$, then $v_i - v_{i+1} - ... - v_{j-1} - v_j$ with $v_i - v_j$ form a cycle with length greater than 3. However, since the graph is chordal, we must have a chord $v_k - v_l$ with $i \leq k, l \leq j and l \geq k + 2$ and $l - k < j - i$, which is contrary to our assumption. □

Lemma 3 is useful for finding chordless path, since checking whether a path is triangle-free is much easier. Another important result is,

Lemma 4. *Let ρ be a cycle with length greater than 3 in a given chordal graph, and X be a vertex on ρ. If the two vertices adjacent to X on ρ are not adjacent to each other, then ρ has a chord of which X is an endpoint.*

Proof. Let v_1 and v_2 be two vertices adjacent to X in ρ. Suppose that ρ doesn't have a chord of which X is an endpoint. Since ρ has length greater than 3, it must have a chord. Clearly, any chord of ρ separates ρ into two sub-cycles, and by our assumptions, it is easy to check that at least one sub-cycle contains X, v_1 and v_2. If this sub-cycle still has a chord, then we can construct another cycle containing X, v_1 and v_2 but has shorter length. Finally, we will have a cycle containing X, v_1 and v_2 but has no chord. Since v_1 and v_2 are not adjacent, the length of this cycle should be greater than 3, Which is contradicted to the definition of chordal graph. □

A chordal graph can be turned into a directed graph by orienting its edges. If the resulting directed graph of an orientation is a DAG without v-structure, then we call this orientation a v-structure-free acyclic orientation. Any v-structure-free acyclic orientation of a connected chordal graph has a unique source, i.e. a vertex which has no parent. Conversely, any vertex in a connected chordal graph can be the unique source in some v-structure-free acyclic orientation. The following lemma is useful.

Lemma 5. *Let $\mathcal{C}$ be a connected chordal graph. For any chordless path π between two distinct vertices X and Y in $\mathcal{C}$, there is a v-structure-free acyclic orientation of $\mathcal{C}$ such that π is a completely directed path from X to Y.*

Proof. Let $\pi = (X = X_0, X_1, ..., X_N = Y)$ be a chordless path from X to Y. Consider a v-structure-free acyclic orientation of $\mathcal{C}$ whose unique source is X. Since X is the source, the edge $X - X_1$ should be oriented as $X \to X_1$. Assume that the subpath $\pi(X, X_{l-1})$ has been oriented as a completely directed path from X to X_{l-1}, then $X_{l-1} - X_l$ should be oriented as $X_{l-1} \to X_l$. Otherwise, $X_{l-2} \to X_{l-1} \leftarrow X_l$ forms a v-structure since by the chordless assumption X_{l-2} is not adjacent to X_l. By induction, π is a completely directed path from X to Y. □

Given a CPDAG $\mathcal{G}^*$, it can be shown that $\mathcal{G}^*$ is a chain graph, which means the undirected subgraph $\mathcal{G}^*_u$ of $\mathcal{G}^*$, i.e. a undirected graph obtained by deleting all the directed edges of $\mathcal{G}^*$, is the union of disjoint chordal graphs, and there is no partially directed cycle in the graph. [MAATHUIS, KALISCH, BUHLMANN] pointed out that any v-structure-free acyclic orientation of the edges in $\mathcal{G}^*_u$ corresponds to a DAG in the equivalence class represented by $\mathcal{G}^*$. and such an orientation can be considered separately for each of the disjoint chordal graphs. Moreover, they proved that,

Lemma 6. *The orientations of the undirected edges adjacent to X are valid if and only if these new directed edges do not introduce v-structures.*

Here, orientations of some edges are called valid if there is a v-structure-free acyclic orientation of $\mathcal{C}$ such that the the directions of these edges in the v-structure-free acyclic orientation coincide with the orientations.

B.1 Proof of Theorem 1

With the help of Lemma 5, we can prove Theorem 1.

Proof. Suppose there is a partially directed path from X to Y in $\mathcal{G}^*$. Without loss of generality, we can assume it has the following form: $X = X_0^1 - X_0^2 - ... - X_0^{k_0} \to X_1^1 - X_1^2 - ... - X_1^{k_1} \to ... \to X_n^1 - X_n^2 - ... - X_n^{k_n} = Y$. Since $X_1^1, X_1^2, ..., X_1^{k_1}$ are in the same chain component, $X_0^{k_0} \to X_1^1$ implies $X_0^{k_0} \to X_1^{k_1}$. Similarly, we have $X_i^{k_i} \to X_{i+1}^{k_{i+1}}$ for $i = 0, 1, 2, ..., n-1$. Hence $X_0^{k_0} \to X_1^{k_1} \to ... \to X_n^{k_n} = Y$ is a completely directed path from $X_0^{k_0}$ to Y. This means $X_0^{k_0}$ is an ancestor of Y in every DAG represented by $\mathcal{G}^*$. On the other hand, Lemma 5 indicates there is a v-structure-free acyclic orientation of the chain component containing X such that X is an ancestor of $X_0^{k_0}$. Combining these two observations we can conclude that X is a potential cause of Y. Conversely, if X is a potential cause of Y, then by the definition of CPDAG and potential cause, we can easily obtain the desired results. □

B.2 Proof of Corollary 1

Proof. According to the definition of partially directed path, an undirected path is also partially directed, hence if X and Y are in the same chain component, they are potential causes of each other. Conversely, if X and Y are potential causes of each other, then by Theorem 1, there is a partially directed path from X to Y as well as a partially directed path from Y to X. Clearly, none of these two paths contains a directed edge, otherwise, a partially directed cycle would occur. Therefore, X and Y are connected by an undirected path, which means they are in the same chain component. □

B.3 Proof of Corollary 2

This corollary is a trivial conclusion hence we omit the proof here.

B.4 Proof of Theorem 2

Proof. It is easy to verify that statement (1) is equivalent to statement (3), thus we only prove (2) ⇔ (3). The necessity is obvious, so we only need to prove the sufficiency. There is no loss of generality in assuming the path has the following form: $X = X_0 \to X_1^1 - X_1^2 - ... - X_1^{k_1} \to ... \to X_n^1 - X_n^2 - ... - X_n^{k_n} = Y$. Similar to the proof of Theorem 1, we have $X \to X_1^{k_1}$ and $X_i^{k_i} \to X_{i+1}^{k_{i+1}}$ for $i = 1, 2, ..., n-1$. Hence, $\rho = (X, X_1^{k_1}, X_2^{k_2}, ..., X_n^{k_n} = Y)$ is a completely directed path from X to Y in $\mathcal{G}^*$, which means ρ is also a directed path in every DAG represented by $\mathcal{G}^*$. □

B.5 Proof of Theorem 3

The proof of Theorem 3 is quiet complicated. First, we show that if X is a semi-invariant cause of Y, then at least two distinct variables in the chain component containing X are completely invariant cause of Y.

Lemma 7. *Let $\mathcal{G}^*$ be a CPDAG. X and Y are two distinct nodes that belong to different chain components. If X is the only completely invariant cause of Y in the chain component to which X belongs, then this chain component doesn't contain any other invariant cause of Y.*

Proof. Let Z be a vertex in the chain component containing X, then every partially directed path between Z and Y passes through X. Since there is a v-structure-free orientation of the chain component whose unique source is X, there is a DAG in the Markov equivalence class represented by $\mathcal{G}^*$ such that none of the vertex except X in the chain component is an ancestor of Y. □

With the help of Lemma 7, we can prove,

Lemma 8. *Let X, Z_1, Z_2 ... Z_n be distinct vertices in a given chordal graph $\mathcal{C}$. The following statements are equivalent.*

(1) $X \in an(\{Z_1, Z_2, ..., Z_n\})$ for every v-structure-free orientation of $\mathcal{C}$;

(2) there exists two distinct nodes Z_i, Z_j such that they are connected by a chordless path through X;

(3) the subgraph over the critical set C of X with respect of $\{Z_1, Z_2, ..., Z_n\}$ in $\mathcal{C}$ is neither empty nor complete.

Proof. It is obvious that we only need to consider the case where X, Z_1, Z_2 ... Z_n are in the same connected component of $\mathcal{C}$.

$(2) \Rightarrow (1)$. If there exists two nodes Z_i, Z_j such that there is a chordless path π between Z_i and Z_j on which X is a intermediate node, then the two vertices adjacent to X on the path, say V_1 and V_2, are not adjacent to each other. Hence, in any v-structure-free acyclic orientation of $\mathcal{C}$, $V_1 \to X$ and $X \leftarrow V_2$ can't exist simultaneously. Denote $\pi = (Z_i, ..., V_1, X, V_2, ..., Z_j)$, if $X \to V_1$, then X is an ancestor of Z_i since the subpath of a chordless path is still chordless. Similarly, if $X \to V_2$, then X is an ancestor of Z_j. Thus, $X \in an(Z_i, Z_j) \subseteq an(Z_1, Z_2, ..., Z_n)$, which completes the proof of $(2) \Rightarrow (1)$.

$(1) \Rightarrow (3)$. Clearly, C is not empty, and by Corollary 2, $n \geq 2$. If C induces a complete subgraph of $\mathcal{C}$, then by Lemma 6, there is a v-structure-free acyclic orientation of $\mathcal{C}$ such that every vertex in C is a parent of X while every vertex in $adj(X) \setminus C$ is a child of X. We will prove that such an orientation leads to a DAG where $X \notin an(Z_1, Z_2, ..., Z_n)$. In fact, if there is a completely directed path $\pi = (X = v_0, v_1, v_2, ..., v_k = Z_i)$ from X to Z_i, then $v_1 \notin C$, which means $k \geq 2$ and π is not chordless. Let $v_p - v_q, q \geq p+2$ be a chord on π. By acyclicity assumption, v_p must point at v_q. Therefore, $\pi(X, v_p)$, $v_p \to v_q$ and $\pi(v_q, Z_i)$ form a completely directed path from X to Z_i. If this path is not chordless, we can find another chord and construct a shorter path which is completely directed from X to Z_i. Finally, we will have a completely directed chordless path π' from X to Z_i. However, the vertex adjacent to X on π' is in critical set, thus the orientation of π' contradicts the assumption that every vertex in C is a parent of X.

$(3) \Rightarrow (2)$. If C doesn't induce a complete subgraph of $\mathcal{C}$, then $\exists C_i, C_j \in C$, C_i and C_j are not adjacent. By the definition of critical set, we have two chordless paths, say π_i and π_j, which are from X to Z_p and Z_q respectively, such that C_i is adjacent to X on π_i and C_j is adjacent to X on π_j. We claim that π_i and π_j have no common vertex except for X. If not, let A be the common vertex such that among all the common vertices, A is closest to X on π_i, then the subpath $\pi_i(X, A)$ has no common vertex with subpath $\pi_j(X, A)$ other than X and A. It is simple to check that $\pi_i(X, A)$ and $\pi_j(X, A)$ form a cycle with length greater than 3, and C_i, X, C_j are three consecutive vertices on this cycle. Since C_i and C_j are not adjacent, by Lemma 4, we have a chord connecting X and some node w on this cycle. Notice that w is either a node on $\pi_i(X, A)$ or a node on $\pi_j(X, A)$, thus the chord $X - w$ is either on π_i or on π_j. This is contradicted to our assumptions. Since π_i and π_j have no common vertex except for X, Z_p and Z_q are distinct. Therefore, π_i and π_j constitute a path π from Z_p to Z_q which passes through X. As π_i and π_j are chordless, they are triangle-free. It follows from C_i and C_j are not adjacent that π is also triangle-free. The desired result comes from Lemma 3. $\square$

Finally, we can prove our main result.

Proof of Theorem 3. $(3) \Rightarrow (2)$ and $(2) \Rightarrow (1)$ can be easily derived from Lemma 8 so we omit the proof here. For $(1) \Rightarrow (3)$, let Z be the set of all completely invariant causes of Y. By Corollary 2 and Lemma 7, X and Y belong to different chain components and $Z \neq \emptyset$. Suppose (3) doesn't hold, then by Lemma 8 X is not an invariant cause of Z, which means there is a DAG $\mathcal{G}$ in the markov equivalence class such that $X \notin an(Z)$. Since X is a semi-invariant cause of Y, it follows that there is a completely directed path from X to Y in $\mathcal{G}$. Let π be such a path, then none of the vertices on π is in Z. Recall that X and Y belong to different chain components, hence by the definition of CPDAG there are two consecutive vertices on π such that the edge between them is directed in $\mathcal{G}^*$. Let v_1 and v_2 be such two vertices with $v_1 \to v_2$ and $\pi(X, v_1)$ corresponds to an undirected path in $\mathcal{G}^*$, then by Theorem 2, v_1 is a completely invariant cause of Y. However, $v_1 \notin Z$, which is contrary to the definition of Z. $\square$

B.6 Proof of Theorem 4

Proof. First we prove statement (1). Suppose X is not a potential cause of Y, that is, X has no causal effect on Y, then for every DAG $\mathcal{G}$ in the Markov equivalence class represented by $\mathcal{G}^*$, Y is a non-descendent of X. Therefore, $X \perp\!\!\!\perp Y | pa(X)$ by Markov Property. On the other hand, if X is a potential cause of Y, then by definition there is a DAG $\mathcal{G}$ in the Markov equivalence class represented by $\mathcal{G}^*$ in which X is an ancestor of Y. Assume π is the directed path from X to Y in $\mathcal{G}$. Since every vertex on π is a non-collider and none of them is in $pa(X)$, $X \not\perp\!\!\!\perp Y | pa(X)$.

Statement (2) is simply a restatement of statement (1), thus we omit the proof.

Next we prove statement (3). If X is a completely invariant cause of Y, then clearly $X \not\perp\!\!\!\perp Y | pa(X)$. By Theorem 2, let π be the directed path from X to Y in $\mathcal{G}^*$, then for any DAG $\mathcal{G}$ in the Markov equivalence class represented by $\mathcal{G}^*$, π is directed, which means π has no collider. However, none of the vertices on π is a member of $pa(X)$ or $sib(X)$, since otherwise, a directed cycle or a partially directed cycle would occur in $\mathcal{G}^*$. Therefore, π is active given $pa(X) \cup sib(X)$, which means $X \not\!\perp\!\!\!\perp Y|pa(X) \cup sib(X)$. Conversely, suppose X is not a completely invariant cause of Y. Since $X \not\!\perp\!\!\!\perp Y|pa(X)$ implies X is a potential cause of Y, X must be a semi-invariant or non-invariant cause of Y. In the following, we will prove in both cases $X \perp\!\!\!\perp Y|pa(X) \cup sib(X)$ holds. By Lemma 6, there is a DAG $\mathcal{G}$ in the Markov equivalence class represented by $\mathcal{G}^*$ such that $ch_{\mathcal{G}}(X) = sib(X) \cup ch(X)$ and $pa_{\mathcal{G}}(X) = pa(X)$. Here, $ch_{\mathcal{G}}(X)$ and $pa_{\mathcal{G}}(X)$ means the children set and parents set of X in $\mathcal{G}$, respectively. Consider a path π from X to Y in $\mathcal{G}$. Without loss of generality, we can assume $\pi = (X, V_1, ..., V_n, Y)$. If $V_1 \in pa_{\mathcal{G}}(X)$, then π is blocked by $pa(X) \cup sib(X)$ since V_1 cannot be a collider on π. If $V_1 \in ch(X)$, then π is not directed, since otherwise, the corresponding path in $\mathcal{G}^*$ would be a a partially directed path from X to Y where the node adjacent to X is a child of X. Therefore, there must be a collider on π. Let V_i is the collider nearest to X. If $V_i \in an(pa(X) \cup sib(X))$, there exists a partially directed cycle in $\mathcal{G}^*$, which is impossible. Thus $V_i \notin an(pa(X) \cup sib(X))$, and π is blocked by $pa(X) \cup sib(X)$. Finally, in the case where $V_1 \in sib(X)$, if V_1 is a non-collider, π is clearly blocked by $pa(X) \cup sib(X)$. If V_1 is a collider, then V_2 is adjacent to X, which means $V_2 \notin ch(X)$, since otherwise, both $X \to V_2 \to V_1 - X$ and $X \to V_2 - V_1 - X$ are partially directed cycles. However, V_2 is a non-collider on π and $V_2 \notin ch(X)$ implies $V_2 \in pa(X) \cup sib(X)$. Hence, π is blocked by $pa(X) \cup sib(X)$. This completes the proof of statement (3).

Again, Statement (4) is simply a restatement of statement (2). We omit the proof here. □

B.7 Proof of Lemma 2

Proof. The construction in A.1 shows that each element in F represents a chordless path starting with X, therefore, to prove the correctness of Algorithm 4, it suffices to show that there is a unique quadruple in F which each corresponds to the given chordless path.

Let $\pi = (X, V_1, V_2, ..., V_m)$ be a chordless path. During the first loop, $(X, \emptyset, 0, 1)$ is added into F, and $(V_1, X, 1)$ is added into S since X and V_2 are not adjacent. Since S is a queue, there exists n_1 such that $(V_1, X, 1)$ will be took out from S during the n_1-th loop. Thus, $(V_1, X, 1, n_1)$ will be added into F. For convenience, we set $V_0 = X, V-1 = \emptyset, n_0 = 1, n_{-1} = 0$. Suppose $(V_i, V_{i-1}, n_{i-1}, n_i), i = 0, 1, ..., m-1$ are all in F, then during the n_{m-1}-th loop, $(V_{m-1}, V_{m-2}, n_{m-2})$ was took out from S, and (V_m, V_{m-1}, n_{m-1}) was added into S. Therefore, there exists n_m such that (V_m, V_{m-1}, n_{m-1}) will be took out from S during the n_m-th loop and $(V_m, V_{m-1}, n_{m-1}, n_m)$ will be added into F. It can be checked that $(V_m, V_{m-1}, n_{m-1}, n_m)$ corresponds to the chordless path π, which prove the existence.

Next we will prove the uniqueness. Suppose that $(V_m, V_{m-1}, s_{m-1}, s_m)$ and $(V_m, V_{m-1}, t_{m-1}, t_m)$ are two different quadruples in F which both correspond to π. By A.1 , the construction procedure of the path will visit a sequence of quadruples in F. Let $(V_i, V_{i-1}, s_{i-1}, s_i)$ and $(V_i, V_{i-1}, t_{i-1}, t_i)$ $i = 0, 1, ..., m$ are two sequences of quadruples which the procedure visited, starting from $(V_m, V_{m-1}, s_{m-1}, s_m)$ and $(V_m, V_{m-1}, t_{m-1}, t_m)$, respectively. Since the forth element of a quadruple is the number of loops when the quadruple enters the queue, $s_m \neq t_m$. If $s_{m-1} = t_{m-1}$, we have $(V_{m-1}, V_{m-2}, s_{m-2}, s_{m-1}) = (V_{m-1}, V_{m-2}, t_{m-2}, t_{m-1})$, which means $(V_m, V_{m-1}, s_{m-1}$ is added to S twice during the s_{m-1}-th loop. This is contrarary to Algorithm 4, hence $s_{m-1} \neq t_{m-1}$. By induction, $s_i \neq t_i$ for $i = 0, 1, ...m$. However, $s_0 = t_0 = 1$. This completes the proof of uniqueness. □

B.8 Proof of Lemma 1

The proof follows similar arguments to the proof of Lemma 2.

B.9 Proof of Theorem 5

The proof follows from Theorem 1 to Theorem 4, as well as Lemma 1.