Luca Weihs*, Bill Robinson, Emilie Dufresne, Jennifer Kenkel, Kaie Kubjas,
Reginald McGee II, Nhan Nguyen, Elina Robeva and Mathias Drton

Determinantal Generalizations of Instrumental Variables

Abstract: Linear structural equation models relate the components of a random vector using linear inter-dependencies and Gaussian noise. Each such model can be naturally associated with a mixed graph whose vertices correspond to the components of the random vector. The graph contains directed edges that represent the linear relationships between components, and bidirected edges that encode unobserved confounding. We study the problem of generic identifiability, that is, whether a generic choice of linear and confounding effects can be uniquely recovered from the joint covariance matrix of the observed random vector. An existing combinatorial criterion for establishing generic identifiability is the half-trek criterion (HTC), which uses the existence of trek systems in the mixed graph to iteratively discover generically invertible linear equation systems in polynomial time. By focusing on edges one at a time, we establish new sufficient and new necessary conditions for generic identifiability of edge effects extending those of the HTC. In particular, we show how edge coefficients can be recovered as quotients of subdeterminants of the covariance matrix, which constitutes a determinantal generalization of formulas obtained when using instrumental variables for identification. While our results do not completely close the gap between existing sufficient and necessary conditions we find, empirically, that our results allow us to prove the generic identifiability of many more mixed graphs than the prior state-of-the-art.

Keywords: trek separation, half-trek criterion, structural equation models, identifiability, generic identifiability

1 Introduction

In a *linear structural equation model* (L-SEM) the joint distribution of a random vector $X = (X_1, \ldots, X_n)^T$ obeys noisy linear interdependencies. These interdependencies can be expressed with a matrix equation of the form

$$X = \lambda_0 + \Lambda^T X + \epsilon, \tag{1}$$

where $\Lambda = (\lambda_{vw}) \in \mathbb{R}^{n \times n}$ and $\lambda_0 = (\lambda_{01}, \ldots, \lambda_{0n})^T \in \mathbb{R}^n$ are unknown parameters, and $\epsilon = (\epsilon_1, \ldots, \epsilon_n)^T$ is a random vector of error terms with positive definite covariance matrix $\Omega = (\omega_{vw})$. Then X has mean vector $(I - \Lambda)^{-T} \lambda_0$ and covariance matrix

$$\phi(\Lambda, \Omega) := (I - \Lambda)^{-T} \Omega (I - \Lambda)^{-1} = \Sigma \tag{2}$$

*Corresponding author: Luca Weihs, Statistics, University of Washington, Box 354322, Seattle, USA, E-mail: lucaw@uw.edu
Bill Robinson, Mathematics, Denison University, Granville, Ohio, USA, E-mail: robinsonw@denison.edu
Emilie Dufresne, University of Nottingham, Nottingham, UK, E-mail: emilie.dufresne@nottingham.ac.uk
Jennifer Kenkel, Mathematics, University of Utah, Salt Lake City, USA, E-mail: kenkel@math.utah.edu
Kaie Kubjas, Mathematics and Systems Analysis, Aalto University, Espoo, Finland, E-mail: kaie.kubjas@aalto.fi
Reginald McGee II, Mathematical Biosciences Institute, Columbus, USA, E-mail: mcgee.278@mbi.osu.ed
Nhan Nguyen, Department of Mathematics, University of Montana, Missoula, USA, E-mail: nhan.nguyen@umontana.edu
Elina Robeva, Department of Mathematics, Massachusetts Institute of Technology, Cambridge, USA, E-mail: erobeva@mit.edu
Mathias Drton, Statistics, University of Washington, Seattle, USA, E-mail: md5@uw.edu

where I is the $n \times n$ identity matrix. L-SEMs have been widely applied in a variety of settings due to the clear causal interpretation of their parameters [1–3].

Following an approach that dates back to Wright [4, 5], we may view Λ and Ω as (weighted) adjacency matrices corresponding to directed and bidirected graphs, respectively. This yields a natural correspondence between L-SEMs and *mixed graphs*, that is, graphs with both directed edges, $v \to w$, and bidirected edges, $v \leftrightarrow w$. More precisely, the mixed graph G is associated to the L-SEM in which λ_{vw} is assumed to be zero if $v \to w \notin G$ and, similarly, $\omega_{vw} = 0$ when $v \leftrightarrow w \notin G$. We write ϕ_G for the map obtained by restricting the map φ from (2) to pairs (Λ, Ω) that satisfy the conditions encoded by the graph G. We note that mixed graphs used to represent L-SEMs are often also called *path diagrams*.

Example 1.1. The mixed graph in Figure 1 corresponds to the well-known instrumental variable model [6]. In equations, this model asserts that

$$X_1 = \lambda_{01} + \epsilon_1, \quad X_2 = \lambda_{02} + \lambda_{12}X_1 + \epsilon_2, \text{ and } \quad X_3 = \lambda_{03} + \lambda_{23}X_2 + \epsilon_3,$$

where ϵ has 0 mean and covariance matrix

$$\Omega = \begin{pmatrix} \omega_{11} & 0 & 0 \\ 0 & \omega_{22} & \omega_{23} \\ 0 & \omega_{23} & \omega_{33} \end{pmatrix}.$$

In this model, the random vector $X = (X_1, X_2, X_3)$ has covariance matrix

$$\Sigma = \begin{pmatrix} 1 & -\lambda_{12} & 0 \\ 0 & 1 & -\lambda_{23} \\ 0 & 0 & 1 \end{pmatrix}^{-T} \begin{pmatrix} \omega_{11} & 0 & 0 \\ 0 & \omega_{22} & \omega_{23} \\ 0 & \omega_{23} & \omega_{33} \end{pmatrix} \begin{pmatrix} 1 & -\lambda_{12} & 0 \\ 0 & 1 & -\lambda_{23} \\ 0 & 0 & 1 \end{pmatrix}^{-1}$$

$$= \begin{pmatrix} \omega_{11} & \lambda_{12}\omega_{11} & \lambda_{12}\lambda_{23}\omega_{11} \\ \lambda_{12}\omega_{11} & \omega_{11}\lambda_{12}^2 + \omega_{22} & \lambda_{23}\omega_{11}\lambda_{12}^2 + \lambda_{23}\omega_{22} + \omega_{23} \\ \lambda_{12}\lambda_{23}\omega_{11} & \lambda_{23}\omega_{11}\lambda_{12}^2 + \lambda_{23}\omega_{22} + \omega_{23} & \omega_{33} + 2\omega_{23}\lambda_{23} + \lambda_{23}^2\sigma_{22} \end{pmatrix}.$$

A first question that arises when specifying an L-SEM via a mixed graph G is whether the map ϕ_G is injective, that is, whether any (Λ, Ω) in the domain of ϕ_G can be uniquely recovered from the covariance matrix $\phi_G(\Lambda, \Omega)$. When this injectivity holds we say that the model and also simply the graph G is *globally identifiable*. Whether or not global identifiability holds can be decided in polynomial time [7–9]. However, in many cases global identifiability is too strong a condition. Indeed, the canonical instrumental variables model is not globally identifiable.

We will be instead interested in *generic identifiability*, that is, whether (Λ, Ω) can be recovered from $\phi_G(\Lambda, \Omega)$ with probability 1 when choosing (Λ, Ω) from any continuous distribution on the domain of ϕ_G. A current state-of-the-art, polynomial time verifiable, criterion for checking generic identifiability of a given mixed graph is the half-trek criterion (HTC) of Foygel et al. [10], with generalizations by [11–13]. The sufficient condition that is part of the HTC operates by iteratively discovering invertible linear equation systems in the Λ parameters which it uses to prove generic identifiability. A necessary condition given by the HTC detects cases in which the Jacobian matrix of ϕ_G fails to attain full column rank which implies that the parameterization ϕ_G is generically infinite-to-one. However, there remain a considerable number of cases in which the HTC remains inconclusive, that is, the graph satisfies the necessary but not the sufficient condition for generic identifiability.

Figure 1: The mixed graph for the instrumental variable model.

We extend the applicability of the HTC in two ways. First, we show how the theorems on trek separation in [14] can be used to discover determinantal relations that in turn can be used to prove the generic identifiability of individual edge coefficients in L-SEMs. This method generalizes the use of conditional independence in known instrumental variable techniques; compare e.g. [15]. Once we have shown that individual edges are generically identifiable with this new method, it would be ideal if identified edges could be integrated into the equation systems discovered by the HTC to prove that even more edges are generically identifiable. Unfortunately, the HTC is not well suited to integrate single edge identifications as it operates simultaneously on all edges incoming to a given node. Our second contribution resolves this issue by providing an *edgewise* half-trek criterion which operates on subsets of a node's parents, rather than all parents at once. This edgewise criterion often identifies many more coefficients than the usual HTC. We note that, in the process of preparing this manuscript we discovered independent work of Chen [16]; some of our results can be seen as a generalization of results in his work.

The rest of this paper is organized as follows. In Section 2, we give a brief overview of the necessary background on mixed graphs, L-SEMs, and the half-trek criterion. In Section 3, we show how trek-separation allows the generic identification of edge coefficients as quotients of subdeterminants. We introduce the edgewise half-trek criterion in Section 4 and we discuss necessary conditions for the generic identifiability of edge coefficients in Section 5. Computational experiments showing the applicability of our sufficient conditions follow in Section 6, and we finish with a brief conclusion in Section 7. Some longer proofs are deferred to the appendix.

2 Preliminaries

We assume some familiarity with the graphical representation of structural equation models and only give a brief overview of our objects of study. A more in-depth introduction can be found, for example, in [2] or, with a focus on the linear case considered here, in [17].

2.1 Mixed graphs and covariance matrices

Nonzero covariances in an L-SEM may arise through direct or through confounding effects. Mixed graphs with two types of edges have been used to represent these two sources of dependences.

Definition 2.1 (Mixed Graph). A *mixed graph* on n vertices is a triple $G = (V, D, B)$ where $V = \{1, \ldots, n\}$ is the vertex set, $D \subset V \times V$ are the directed edges, and $B \subset V \times V$ are the bidirected edges. We require that there be no self-loops, so $(v, v) \notin D, B$ for all $v \in V$. If $(v, w) \in D$, we will write $v \to w \in G$ and if $(v, w) \in B$, we will write $v \leftrightarrow w \in G$. As bidirected edges are symmetric we will also require that B is symmetric, so that $(v, w) \in B \iff (w, v) \in B$.

Let v and w be two vertices of a mixed graph $G = (V, D, B)$. A *path* from v to w is any sequence of edges from D or B beginning at v and ending at w. Here, we allow that directed edges be traversed against their natural direction (i.e., from head to tail). We also allow repeated vertices on a path. Sometimes, such paths are referred to as walks or also semi-walks. A path from v to w is *directed* if all of its edges are directed and point in the same direction, away from v and towards w.

Definition 2.2 (Treks and half-treks). (a) A path π from a *source* v to a *target* w is a *trek* if it has no colliding arrowheads, that is, π is of the form

$$v_l^L \leftarrow v_{l-1}^L \leftarrow \cdots \leftarrow v_0^L \longleftrightarrow v_0^R \to v_1^R \to \cdots \to v_{r-1}^R \to v_r^R \quad \text{or}$$
$$v_l^L \leftarrow v_{l-1}^L \leftarrow \cdots \leftarrow v_1^L \leftarrow v^T \to v_1^R \to \cdots \to v_{r-1}^R \to v_r^R,$$

where $v_l^L = v$, $v_r^R = w$, and v^T is the *top* node. Each trek π has a left-hand side Left(π) and a right-hand side Right(π). In the former case, Left(π) = $\{v_0^L, \ldots, v_l^L\}$ and Right(π) = $\{v_0^R, \ldots, v_r^R\}$. In the latter case, Left(π) = $\{v^T, v_1^L, \ldots, v_l^L\}$ and Right(π) = $\{v^T, v_1^R, \ldots, v_r^R\}$, with v^T a part of both sides.

(b) A trek π is a *half-trek* if |Left(π)| = 1. In this case π is of the form

$$v_0^L \longleftrightarrow v_0^R \to v_1^R \to \cdots \to v_{r-1}^R \to v_r^R \quad \text{or} \quad v^T \to v_1^R \to \cdots \to v_{r-1}^R \to v_r^R.$$

In particular, a half-trek from v to w is a trek from v to w which is either empty, begins with a bidirected edge, or begins with a directed edge pointing away from v.

Some terminology is needed to reference the local neighborhood structure of a vertex v. For the directed part (V, D), it is standard to define the set of *parents* and the set of *descendents* of v as

$$\text{pa}(v) = \{w \in V : w \to v \in G\},$$
$$\text{des}(v) = \{w \in V : \exists \text{ a non-empty directed path from } v \text{ to } w \text{ in } G\},$$

respectively. The nodes incident to a bidirected edge can be thought of as having a common (latent) parent and thus we refer to the bidirected neighbors as *siblings* and define

$$\text{sib}(v) = \{w \in V : w \leftrightarrow v \in G\}.$$

Finally, we denote the sets of nodes that are *trek reachable* or *half-trek reachable* from v by

$$\text{tr}(v) = \{w \in V : \exists \text{ a non-empty trek from } v \text{ to } w \text{ in } G\},$$
$$\text{htr}(v) = \{w \in V : \exists \text{ a non-empty half-trek from } v \text{ to } w \text{ in } G\}.$$

Two sets of matrices may be associated with a given mixed graph $G = (V, D, B)$. First, $\mathbb{R}_{\text{reg}}^D$ is the set of real $n \times n$ matrices $\Lambda = (\lambda_{vw})$ with support D, i.e., those matrices Λ with $\lambda_{vw} \neq 0$ implying $v \to w \in G$ and for which $I - \Lambda$ invertible. Second, $PD(B)$ is the set of positive definite matrices with support B, i.e., if $v \neq w$, then $\omega_{vw} \neq 0$ implies $v \leftrightarrow w \in G$. Based on (2), the distributions in the L-SEM given by G have a covariance matrix Σ that is parameterized by the map

$$\phi_G : (\Lambda, \Omega) \mapsto (I - \Lambda)^{-T} \Omega (I - \Lambda)^{-1} \tag{3}$$

with domain $\Theta := \mathbb{R}_{\text{reg}}^D \times PD(B)$.

Remark 2.3. Our focus is solely on covariance matrices. Indeed, in the traditional case where the errors ϵ in (1) follow a multivariate normal distribution the covariance matrix contains all available information about the parameters (Λ, Ω).

Subsequently, the matrices Λ, Ω and Σ will also be regarded as matrices of indeterminants. The entries of $(I - \Lambda)^{-1} = I + \sum_{k=1}^{\infty} \Lambda^k$ may then be interpreted as formal power series. Let Λ and Ω be matrices of indeterminants with zero pattern corresponding to G. Then $\Sigma = \phi_G(\Lambda, \Omega)$ has entries that are formal power series whose form is described by the Trek Rule of Wright [4], see also Spirtes, Glymour, and Scheines [3]. The Trek rule states that for every $v, w \in V$ the corresponding entry of $\phi_G(\Lambda, \Omega)$ is the sum of all trek monomials corresponding to treks from v to w.

Definition 2.4 (Trek Monomial). Let $v, w \in V$ be two, not necessarily distinct, vertices, and let $\mathscr{T}(v, w)$ be the set of all treks from v to w in G. If $\pi \in \mathscr{T}(v, w)$ contains no bidirected edge and has top node z, its *trek monomial* is defined as

$$\pi(\Lambda, \Omega) = \omega_{zz} \prod_{x \to y \in \pi} \lambda_{xy}.$$

If π contains a bidirected edge connecting $u, z \in V$, then its trek monomial is

$$\pi(\Lambda, \Omega) = \omega_{uz} \prod_{x \to y \in \pi} \lambda_{xy}.$$

Proposition 2.5 (Trek Rule). *The covariance matrix $\Sigma = \phi_G(\Lambda, \Omega)$ corresponding to a mixed graph G satisfies*

$$\Sigma_{vw} = \sum_{\pi \in \mathscr{T}(v,w)} \pi(\Lambda, \Omega), \quad v, w \in V.$$

2.2 Generic identifiability

We now formally introduce our problem of interest and review some of the prior work our results build on. We recall that an *algebraic set* is the zero-set of a collection of polynomials. An algebraic set that is a proper subset of Euclidean space has measure zero; see, e.g., the lemma in [18].

Definition 2.6 (Generic Identifiability). (a) The model given by a mixed graph G is *generically identifiable* if there exists a proper algebraic subset $A \subset \Theta$ such that the fiber $\mathscr{F}(\Lambda, \Omega) := \phi_G^{-1}(\{\phi_G(\Lambda, \Omega)\})$ is a singleton set, that is, it satisfies

$$\mathscr{F}(\Lambda, \Omega) = \{(\Lambda, \Omega)\}$$

for all $(\Lambda, \Omega) \in \Theta \backslash A$. In this case we will say, for simplicity, that G is generically identifiable.

(b) Let $\mathrm{proj}_{v \to w}$ be the projection $(\Lambda, \Omega) \mapsto \lambda_{vw}$ for $v \to w \in G$. We say that the edge coefficient λ_{vw} is *generically identifiable* if there exists a proper algebraic subset $A \subset \Theta$ such that $\mathrm{proj}_{v \to w}(\mathscr{F}(\Lambda, \Omega)) = \{\lambda_{vw}\}$ for all $(\Lambda, \Omega) \in \Theta \backslash A$. In this case, we will say that the edge $v \to w$ is generically identifiable.

In all examples we know of, if generic identifiability holds, then the parameters can in fact be recovered using rational formulas.

Definition 2.7 (Rational Identifiability). (a) A mixed graph G, or rather the model it defines, is *rationally identifiable* if there exists a rational map ψ and a proper algebraic subset $A \subset \Theta$ such that $\psi \circ \phi_G$ is the identity on $\Theta \backslash A$.

(b) An edge $v \to w \in G$, or rather the coefficient λ_{vw}, is *rationally identifiable* if there exists a rational function ψ and a proper algebraic subset $A \subset \Theta$ such that $\psi \circ \phi_G(\Lambda, \Omega) = \lambda_{vw}$ for all $(\Lambda, \Omega) \in \Theta \backslash A$.

We now introduce the half-trek criterion (HTC) from [10]. We generalize this criterion in Section 4.

Definition 2.8 (Trek and Half-Trek Systems). Let $\Pi = \{\pi_1, \ldots, \pi_m\}$ be a collection of treks in G and let S, T be the set of sources and targets of the π_i respectively. Then we say that Π is a *system of treks* from S to T. If each π_i is a half-trek, then Π is a *system of half-treks*. A collection $\Pi = \{\pi_1, \ldots, \pi_m\}$ of treks is said to have *no sided intersection* if

$$\mathrm{Left}(\pi_i) \cap \mathrm{Left}(\pi_j) = \varnothing = \mathrm{Right}(\pi_i) \cap \mathrm{Right}(\pi_j), \ \forall i \neq j.$$

As our focus will be on the identification of individual edges in G we do not state the identifiability result of Foygel et al. [10] in its usual form, instead we present a slightly modified version which is easily seen to be implied by the proof of Theorem 1 in [10].

Definition 2.9. *A set of nodes $Y \subset V$ satisfies the half-trek criterion* with respect to a vertex $v \in V$ if
(i) $|Y| = |\mathrm{pa}(v)|$,
(ii) $Y \cap (\{v\} \cup \mathrm{sib}(v)) = \varnothing$, and
(iii) there is a system of half-treks with no sided intersection from Y to $\mathrm{pa}(v)$.

Theorem 2.10 (HTC-identifiability). *Suppose that in the mixed graph $G = (V, D, B)$ the set $Y \subset V$ satisfies the half-trek criterion with respect to $v \in V$. If all directed edges $u \to y \in G$ with head $y \in htr(v) \cap Y$ are generically (rationally) identifiable, then all directed edges with v as a head are generically (rationally) identifiable.*

The sufficient condition for rational identifiability of G in [10] is obtained through iterative application of Theorem 2.10.

3 Trek separation and identification by ratios of determinants

Let Λ and Ω be matrices of indeterminants corresponding to a mixed graph $G = (V, D, B)$ as specified in Section 2.1. Let $S, T \subset V$, and let $\Sigma_{S,T}$ be the submatrix of $\Sigma = \phi_G(\Lambda, \Omega) \in \mathbb{R}^{n \times n}$ obtained by retaining only the rows and columns indexed by S and T, respectively. The (generic) rank of such a submatrix $\Sigma_{S,T}$ can be completely characterized by considering the trek systems between the vertices in S and T. The formal statement of this result follows.

Definition 3.1 (t-separation). A pair of sets (L, R) with $L, R \subset V$ *t-separates* the sets $S, T \subset V$ if every trek between a vertex $s \in S$ and a vertex $t \in T$ intersects L on the left or R on the right.

In this definition, the symbols L and R are chosen to suggest left and right. Similarly, S and T are chosen to indicate sources and targets, respectively.

Theorem 3.2 ([14], [19]). *Let r be a non-negative integer. The submatrix $\Sigma_{S,T}$ has generic rank $\leq r$ if and only if there exist sets $L, R \subset V$ with $|L| + |R| \leq r$ such that (L, R) t-separates S and T.*

Theorem 2.7 from [14] established this result for acyclic mixed graphs while [19] extended the result to all mixed graphs and even gave an explicit representation of the rational form of the subdeterminant $|\Sigma_{S,T}|$, for $|S| = |T|$. An immediate corollary to the above theorem, considering the proof of Theorem 2.17 in [14], rephrases its statement in terms of maximum flows in a special graph. For an introduction to maximum flow, and the well-known Max-flow Min-cut Theorem, see the book by Cormen et al. [20]. Note that standard max-flow min-cut framework does not allow vertices to have maximum capacities or for there to be multiple sources and targets, introducing these modifications is, however, trivial and the resulting theorem is sometimes called the Generalized Max-flow Min-cut Theorem.

Corollary 3.3. *Let $G_{flow} = (V_f, D_f)$ be the directed graph with $V_f = \{1, \dots, n\} \cup \{1', \dots, n'\}$ and D_f containing the following edges:*

$$i \to j \quad if\, j \to i \in G, \tag{4}$$

$$i \to i' \quad for\, all\, i \in V, \tag{5}$$

$$i \to j' \quad if\, i \leftrightarrow j \in G,\ and \tag{6}$$

$$i' \to j' \quad if\, i \to j \in G. \tag{7}$$

Turn G_{flow} into a network by giving all vertices and edges capacity 1. Let $S = \{s_1, \dots, s_k\}, T = \{t_1, \dots, t_m\} \subset V$. Then $\Sigma_{S,T}$ has generic rank r if and only if the max-flow from $s_1, \dots, s_k$ to $t'_1, \dots, t'_m$ in G_{flow} is r.

Proof. Add vertices u, v, with infinite capacity, to the graph G_{flow} along with edges, all with capacity 1, $u \to s_i$, for $1 \leq i \leq k$, and $t'_j \to v$, for $1 \leq j \leq m$. Let L, R be such that they t-separate the sets S, T and $|L| + |R|$ is minimal. By Theorem 3.2, $\Sigma_{S,T}$ has rank $|L| + |R|$ generically. Note that $L \cup R$ gives the minimal size $s - t$ cut (of size $|L| + |R|$). By the (generalized) Max-flow Min-cut theorem the max-flow from u to v is $|L| + |R|$, and it is also the max flow from $s_1, \dots, s_k$ to $t'_1, \dots, t'_m$. Hence $\Sigma_{S,T}$ has generic rank equal to the found max-flow. □

Note that the maximum flow between vertex sets in a graph can be computed in polynomial time. Indeed, in our case, the conditions of Corollary 3.3 can be checked in $O(|V|^2 \max\{m, k\})$ time [20, page 725]. As the
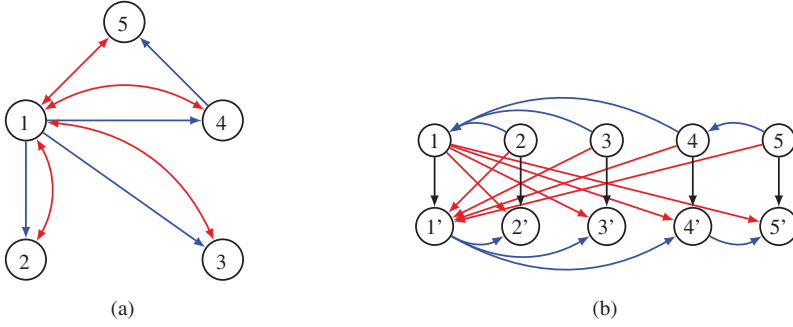
Figure 2: (a) A graph G that is generically identifiable but for which the HTC fails to identify any coefficients. (b) The corresponding flow graph G_{flow}, black edges correspond to (5), red edges to (6), and blue edges to (4) and (7).
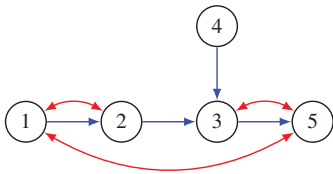

Figure 3: A graph for which Theorem 3.8 can be used to to certify that the edge λ_{12} is identifiable when Theorem 3.5 cannot.

following example shows, Corollary 3.3 can be used to find determinantal constraints on Σ. These constraints can then be leveraged to identify edges in G.

Example 3.4. Consider the mixed graph $G = (V, D, B)$ in Figure 2a, which is taken from Figure 3c in [10]. The corresponding flow network G_{flow} is shown in Figure 2b. From Gröbner basis computations, G is known to be rationally identifiable but the half-trek criterion fails to certify that any edge of G is generically identifiable. Let $S = \{1, 2, 4\}$ and $T = \{1, 3, 5\}$. Corollary 3.3 implies that $\Sigma_{S,T}$ has generically full rank as there is a flow of size 3 from S to $T' = \{1', 3', 5'\}$ in G_{flow}, via the paths $1 \to 3'$, $2 \to 1'$, and $4 \to 5'$. Now suppose that we remove the $4 \to 5$ edge from G, call the resulting graph $\bar{G}$, and let $\bar{\Sigma}$ be the covariance matrix corresponding to $\bar{G}$. Then one may check that the max-flow from S to T' in $\bar{G}_{flow}$ is ≤ 2. Thus $|\bar{\Sigma}_{\{1,2,4\},\{1,3,5\}}| = 0$ where $|\cdot|$ denotes the determinant. Now note that $\lambda_{45}\sigma_{14}$ is the sum of all monomials given by treks from 1 to 5 that end in the edge λ_{45}. Hence, $\sigma_{15} - \lambda_{45}\sigma_{14}$ is obtained by summing over all treks from 1 to 5 that do not end in the edge $4 \to 5$. But in our graph this is just the sum over treks from 1 to 5 that do not use the edge $4 \to 5$ at all. Therefore, $\bar{\sigma}_{15} = \sigma_{15} - \lambda_{45}\sigma_{14}$. Similarly, it is straightforward to check that

$$\bar{\Sigma}_{\{1,2,4\},\{1,3,5\}} = \begin{pmatrix} \sigma_{11} & \sigma_{13} & \sigma_{15} - \lambda_{45}\sigma_{14} \\ \sigma_{21} & \sigma_{23} & \sigma_{25} - \lambda_{45}\sigma_{24} \\ \sigma_{41} & \sigma_{43} & \sigma_{45} - \lambda_{45}\sigma_{44} \end{pmatrix}. \tag{8}$$

By the multilinearity of the determinant, we deduce that

$$0 = |\bar{\Sigma}_{\{1,2,4\},\{1,3,5\}}| = \begin{vmatrix} \sigma_{11} & \sigma_{13} & \sigma_{15} \\ \sigma_{21} & \sigma_{23} & \sigma_{25} \\ \sigma_{41} & \sigma_{43} & \sigma_{45} \end{vmatrix} - \lambda_{45} \begin{vmatrix} \sigma_{11} & \sigma_{13} & \sigma_{14} \\ \sigma_{21} & \sigma_{23} & \sigma_{24} \\ \sigma_{41} & \sigma_{43} & \sigma_{44} \end{vmatrix}$$

$$= |\Sigma_{\{1,2,4\},\{1,3,5\}}| - \lambda_{45}|\Sigma_{\{1,2,4\},\{1,3,4\}}|.$$

Applying Corollary 3.3 a final time, we recognize that $|\Sigma_{\{1,2,4\},\{1,3,4\}}|$ is generically non-zero and, thus, the equation

$$\lambda_{45} = \frac{|\Sigma_{\{1,2,4\},\{1,3,5\}}|}{|\Sigma_{\{1,2,4\},\{1,3,4\}}|}$$

generically and rationally identifies λ_{45}. In this case, the same strategy can be used to identify the edges $1 \to 2$ and $1 \to 3$ (but not $1 \to 4$) in G.

In the above example, there is a correspondence between trek systems in G and trek systems in $\bar{G}$, the graph that has the edge to be identified removed. This allowed us to leverage Corollary 3.3 directly to show that (8) has determinant 0. Such a correspondence cannot always be obtained but exists in the following case.

Theorem 3.5. *Let $G = (V, D, B)$ be a mixed graph. Let $w_0 \to v$ be an edge in G, and suppose that the edges $w_1 \to v, \ldots, w_\ell \to v \in G$ are known to be generically (rationally) identifiable. Let $\bar{G}$ be the subgraph of G with the edges $w_0 \to v, \ldots, w_\ell \to v \in G$ removed. Suppose there are sets $S \subset V\backslash\{v\}$, $T \subset V\backslash\{v, w_0\}$ with $|S| = |T| + 1 = k$ such that:*
(a) $\mathrm{des}(v) \cap (S \cup T \cup \{v\}) = \varnothing$,
(b) *the max-flow from S to $T' \cup \{w_0'\}$ in G_{flow} equals k, and*
(c) *the max-flow from S to $T' \cup \{v'\}$ in $\bar{G}_{flow}$ is smaller than k.*

Then $w_0 \to v$ is generically (rationally) identifiable by the equation

$$\lambda_{w_0 v} = \frac{|\Sigma_{S,T \cup \{v\}}| - \sum_{i=1}^{\ell} \lambda_{w_i v}|\Sigma_{S,T \cup \{w_i\}}|}{|\Sigma_{S,T \cup \{w_0\}}|}. \tag{9}$$

Proof. Let Σ and $\bar{\Sigma}$ be the covariance matrices corresponding to G and $\bar{G}$, respectively. Since $\mathrm{des}(v) \cap (S \cup T \cup \{v\}) = \varnothing$, we have that $\sigma_{st} = \bar{\sigma}_{st}$ for all $s \in S$ and $t \in T$. This holds because if a trek from s to t uses an edge $w_i \to v$ then either $s \in \{v\} \cup \mathrm{des}(v)$ or $t \in \{v\} \cup \mathrm{des}(v)$, violating our assumptions.

Now let $s \in S$ and $0 \le i \le \ell$. Suppose that π is a trek from s to v that uses the edge $w_i \to v$. Then since $s \notin \{v\} \cup \mathrm{des}(v)$ we must have that $w_i \to v$ is used only on the right-hand side of π. With $v \notin \mathrm{des}(v)$ it follows that $w_i \to v$ is the last edge used in the trek because π may only use directed edges after using $w_i \to v$ and must end at v. Hence, all treks from s to v which use $w_i \to v$ must have this edge as their last edge on the right. But $\sigma_{sw_i}\lambda_{w_i v}$ is obtained by summing over all treks from s to v which end in the edge $w_i \to v$ and, thus, $\sigma_{sv} - \sigma_{sw_i}\lambda_{w_i v}$ is the sum of the monomials for all treks from s to v that do not use the $w_i \to v$ edge at all.

As the above argument holds for all $0 \le i \le \ell$, it follows that $\bar{\sigma}_{sv} = \sigma_{sv} - \sum_{i=0}^{k} \sigma_{sw_i}\lambda_{w_i v}$. Since this is true for all $s \in S$ it follows, similarly as in Example 3.4, that

$$|\bar{\Sigma}_{S,T \cup \{v\}}| = |\Sigma_{S,T \cup \{v\}}| - \sum_{i=0}^{k} \lambda_{w_i v}|\Sigma_{S,T \cup \{w_i\}}|.$$

Using assumption (c) and applying Corollary 3.3, we have $|\bar{\Sigma}_{S,T \cup \{v\}}| = 0$. Similarly, by assumption (b), $|\Sigma_{S,T \cup \{w_0\}}| \ne 0$ generically. The desired result follows.   □

Remark 3.6. Theorem 3.5 generalizes the ideas underlying instrumental variable methods such as those discussed in [15]. Indeed, this prior work uses d-separation as opposed to t-separation. D-separation characterizes conditional independence which in the present context corresponds to the vanishing of particular almost principal determinants of the covariance matrix. In contrast, Theorem 3.5 allows us to leverage arbitrary determinantal relations; compare [14]. The graph in Figure 2a is an example in which d-separation and traditional instrumental variable techniques cannot explain the rational identifiability of the coefficient for edge $4 \to 5$.

While assumption (a) in the above Theorem allows for the easy application of Corollary 3.3, this assumption can be relaxed by generalizing one direction of Corollary 3.3. We state this generalization as the following lemma, which is concerned with asymmetric treatment of edges that appear on the left versus right-hand side of treks. The lemma's proof is deferred to Appendix A.

Lemma 3.7. *Let $G = (V, D, B)$ be a mixed graph, and let $\Lambda = (\lambda_{uv})$ and Ω be the matrices of indeterminants corresponding to the directed and the bidirected part of G, respectively. Let $D_L, D_R \subset D$ and define $n \times n$ matrices Λ^L and Λ^R with*

$$\Lambda^L_{uv} = \begin{cases} \lambda_{uv} & \text{if } (u, v) \in D_L, \\ 0 & \text{otherwise,} \end{cases} \quad \text{and}$$

$$\Lambda^R_{uv} = \begin{cases} \lambda_{uv} & \text{if } (u, v) \in D_R, \\ 0 & \text{otherwise.} \end{cases}$$

Define a network $G^*_{flow} = (V^*, D^*)$ with vertex set $V^* = \{1, \dots, n\} \cup \{1', \dots, n'\}$, edge set D^* containing

$$i \to j \;\; \text{if } (j, i) \in D_L, \tag{10}$$
$$i \to i' \;\; \text{for all } i \in V, \tag{11}$$
$$i \to j' \;\; \text{if } (i, j) \in B, \tag{12}$$
$$i' \to j' \;\; \text{if } (i, j) \in D_R, \;\; \text{and} \tag{13}$$

with all edges and vertices having capacity 1. Let $\Gamma = (I - \Lambda^L)^{-T} \Omega (I - \Lambda^R)^{-1}$. Then, for any $S, T \subset V$ with $|S| = |T| = k$, we have that $|\Gamma_{S,T}| = 0$ if the max-flow from S to T' in G^*_{flow} is $< k$.

We may now state our more general result.

Theorem 3.8. *Let $G = (V, D, B)$ be a mixed graph, $w_0 \to v \in G$, and suppose that the edges $w_1 \to v, \dots, w_\ell \to v \in G$ are known to be generically (rationally) identifiable. Recalling Equation (13), let G^*_{flow} be G_{flow} with the edges $w'_0 \to v', \dots, w'_\ell \to v'$ removed. Suppose there are sets $S \subset V$ and $T \subset V \backslash \{v, w_0\}$ such that $|S| = |T| + 1 = k$ and*
(a) $\text{des}(v) \cap (T \cup \{v\}) = \varnothing$,
(b) the max-flow from S to $T' \cup \{w'_0\}$ in G_{flow} equals k, and
(c) the max-flow from S to $T' \cup \{v'\}$ in G^*_{flow} is $< k$.

Then $w_0 \to v$ is rationally identifiable by the equation

$$\lambda_{w_0 v} = \frac{|\Sigma_{S, T \cup \{v\}}| - \sum_{i=1}^{\ell} \lambda_{w_i v} |\Sigma_{S, T \cup \{w_i\}}|}{|\Sigma_{S, T \cup \{w_0\}}|}. \tag{14}$$

Proof. By assumption (b) and Corollary 3.3, $|\Sigma_{S, T \cup \{w_0\}}|$ is generically non-zero. Therefore, equation (14) holds if

$$|\Sigma_{S, T \cup \{v\}}| - \sum_{i=0}^{\ell} \lambda_{w_i v} |\Sigma_{S, T \cup \{w_i\}}| = 0.$$

To show this we note that, by the multilinearity of the determinant, we have

$$|\Sigma_{S, T \cup \{v\}}| - \sum_{i=0}^{\ell} \lambda_{w_i v} |\Sigma_{S, T \cup \{w_i\}}| = \begin{vmatrix} \sigma_{s_1 t_1} & \cdots & \sigma_{s_1 t_{k-1}} & \sigma_{s_1 v} - \sum_{i=0}^{\ell} \lambda_{w_i v} \sigma_{s_1 w_i} \\ \sigma_{s_2 t_1} & \cdots & \sigma_{s_2 t_{k-1}} & \sigma_{s_2 v} - \sum_{i=0}^{\ell} \lambda_{w_i v} \sigma_{s_2 w_i} \\ \vdots & \ddots & \vdots & \vdots \\ \sigma_{s_k t_1} & \cdots & \sigma_{s_k t_{k-1}} & \sigma_{s_k v} - \sum_{i=0}^{\ell} \lambda_{w_i v} \sigma_{s_k w_i} \end{vmatrix}.$$

Write Γ for the matrix that appears on the right-hand side of this equation.

Consider any two indices i and j with $1 \leq i \leq k$ and $1 \leq j \leq k - 1$. If a trek from s_i to t_j uses one of the edges $w_m \rightarrow v$, for $0 \leq m \leq \ell$, on its right-hand side then $t_j \in \text{des}(v)$, a contradiction since $\text{des}(v) \cap T = \varnothing$ by assumption (a). Similarly, since $v \notin \text{des}(v)$ the difference $\sigma_{s_i v} - \sum_{j=0}^{\ell} \lambda_{w_j v} \sigma_{s_i w_j}$ is obtained by summing the monomials for treks between s_i and v which do not use any edge $w_j \rightarrow v$ on their right side. From this we may write

$$\Gamma = ((I - \Lambda)^{-T} \Omega (I - \Lambda')^{-1})_{S, T \cup \{v\}}$$

where Λ' equals Λ but with its (w_j, v), $0 \leq j \leq \ell$, entries set to 0. The fact that $|\Gamma| = 0$ under assumption (c) is the content of Lemma 3.7 (where we take $\Lambda^L = \Lambda$ and $\Lambda^R = \Lambda'$). Given this lemma our desired result then follows. $\square$

Clearly Theorem 3.8 can be applied whenever Theorem 3.5 can. Moreover, as the next example shows, there are cases in which Theorem 3.8 can be used while Theorem 3.5 cannot.

Example 3.9. Let $G = (V, D, B)$ be the mixed graph from Figure 3. Take $S = \{3, 5\}$ and $T = \{4\}$. Then Theorem 3.8 implies that λ_{12} is rationally identifiable. Theorem 3.5 cannot be applied in this case as $S \cap \text{des}(2) \neq \varnothing$.

For a fixed choice of S and T, the conditions (a)-(c) in Theorem 3.8 can be verified in polynomial time. Indeed, conditions (b) and (c) involve only max-flow computations that take $O(|V|^3)$ time in general. Condition (a) can be checked by computing the descendants of v, which can be done with any $O(|D|)$ graph traversal algorithm (e.g., depth first search, see Cormen et al. [20]), and then computing the intersection between the descendants and $T \cup \{v\}$ which can be done in $O(|V| \log |V|)$ time.

In order to apply Theorem 3.8 algorithmically, however, we have to consider all possible subsets $S \subset V$, $T \subset V \backslash \{v, w_0\}$ and check our condition for each pair. Naively done this operation takes exponential time. It remains an interesting problem for further study to determine whether or not the problem of finding suitable sets S and T is NP-hard. We note that a similar problem arises for instrumental variables/d-separation, where [21] were able to give a polynomial time algorithm for finding suitable sets in graphs that are acyclic. Given our results so far we will maintain polynomial time guarantees simply by considering only subsets S, T of bounded size $|S|, |T| \leq m$.

4 Edgewise generic identifiability

While our results from Section 3 can be used together with the HTC there is notable lack of synergy between the two methods as Theorem 3.8 requires that all directed edges incoming to a node be generically identifiable before that node can be used to prove the generic identifiability of other edges. Aiming to strengthen the HTC while allowing it to better use identifications produced by Theorem 3.8, the following theorem establishes a sufficient condition for the generic identifiability of any set of incoming edges to a fixed node. While in the process of preparing this manuscript we discovered the work of Chen [16]; our following theorem can be seen as a generalization of his Theorem 1, see Remark 4.2 for a discussion of the primary difference between our theorem and that in [16].

Theorem 4.1. *Let $G = (V, D, B)$ be a, non-empty, mixed graph and let $v \in V$. Let $W \subset pa(v)$ and suppose there exists $Y \subset V \backslash (\{v\} \cup sib(v))$ with $|Y| = |W|$ such that,*
(i) there exists a half-trek system from Y to W with no sided intersection,
(ii) for every trek π from $y \in Y$ to v we have that either
   (a) π ends with an edge of the form $s \rightarrow v$ where either $s \in W$ or $s \rightarrow v$ is known to be generically (rationally) identifiable, or
   (b) π begins with an edge of the form $y \leftarrow s$ where $s \rightarrow y$ is known to be generically (rationally) identifiable.

Then for each $w \in W$ we have that $w \rightarrow v$ is generically (rationally) identifiable.

Proof. Let (Λ, Ω) be the matrices of indeterminants corresponding to G, and let $\Sigma = (I - \Lambda)^{-T}\Omega(I - \Lambda)^{-1}$ be the covariance matrix. Recall our notation $\mathcal{T}(x, z)$ for the set of treks from x to z in G. By the trek rule (Prop. 2.5), $\Sigma_{xz} = \sum_{\pi \in \mathcal{T}(x,z)} \pi(\Lambda, \Omega)$ is the sum of monomials for treks from v to w.

Recalling that $|Y| = |W|$, enumerate $W = \{w_1, \dots, w_k\}$ and $Y = \{y_1, \dots, y_k\}$. Now, for $1 \le i \le k$, let $H_i \subset D$ be the set of all edges incoming to y_i known to be generically (rationally) identifiable.

Our approach is to build a linear system of k equations in the k unknowns $\lambda_{w_1 v}, \dots, \lambda_{w_k v}$, having a unique solution. Consider the set $\mathcal{T}(y_1, v)$ of all treks between y_1 and v. Because of condition (ii) we have that $y_1 \leftrightarrow v \notin G$ and all treks from y_1 to v either end in a directed edge of the form $s \to v$, with $s \in W$ or $s \to v$ known to be generically identifiable, or must start in a directed edge of the form $y_1 \leftarrow h$ for some $h \in H_1$. Now note that for any $p \in \mathrm{pa}(v)$,

$$\sum_{\pi \in \mathcal{T}(y_1, p)} \pi(\Lambda, \Omega) - \sum_{h \in H_1} \sum_{\pi \in \mathcal{T}(h,p)} \pi(\Lambda, \Omega)\lambda_{py_1} = \Sigma_{y_1 p} - \sum_{h \in H_1} \Sigma_{hp}\lambda_{hy_1}$$

equals the sum of the monomials for all treks from y_1 to p that do not start with a directed edge of the form $y_1 \leftarrow h$ for $h \in H$. Hence we find that the sum of all monomials for treks from y_1 to v that do not start with an edge of the form $y_1 \leftarrow h$ for $h \in H_1$ equals

$$\sum_{p \in \mathrm{pa}(v)} (\Sigma_{y_1 p} - \sum_{h \in H_1} \Sigma_{hp}\lambda_{hy_1})\lambda_{pv}.$$

Now the sum over all treks between y_1 and v that start with an edge of the form $y_1 \leftarrow h$ for $h \in H_1$ is easily seen to be the quantity $\sum_{h \in H_1} \Sigma_{vh}\lambda_{hy_1}$. Thus,

$$\Sigma_{y_1 v} = \sum_{p \in \mathrm{pa}(v)} (\Sigma_{y_1 p} - \sum_{h \in H_1} \Sigma_{hp}\lambda_{hy_1})\lambda_{pv} + \sum_{h \in H_1} \Sigma_{vh}\lambda_{hy_1}$$

Rewriting this we have

$$\sum_{w \in W} (\Sigma_{y_1 w} - \sum_{h \in H_1} \Sigma_{hw}\lambda_{hy_1})\lambda_{wv}$$
$$= \Sigma_{y_1 v} - \sum_{p \in \mathrm{pa}(v) \backslash W} (\Sigma_{y_1 p} - \sum_{h \in H_1} \Sigma_{hp}\lambda_{hy_1})\lambda_{pv} - \sum_{h \in H_1} \Sigma_{vh}\lambda_{hy_1}.$$

Notice that, in the above equation, if $p \in \mathrm{pa}(v) \backslash W$ and $\Sigma_{y_1 p} - \sum_{h \in H_1} \Sigma_{hp}\lambda_{hy_1} \ne 0$ then it must be the case that there is a trek π from y_1 to v ending in the edge $p \to v$ which does not start with an edge of the form $y_1 \leftarrow s$ where $s \to y_1$ is known to be generically identifiable. It then follows, by condition (ii)(a), that since $p \notin W$ we must have that λ_{pv} is known to be generically identifiable. It then follows that the only unknowns quantities (that is, those not assumed to be generically identifiable) in the above displayed equation are the λ_{wv} which appear linearly on the left hand side. Thus we have exhibited one linear equation in the k unknown parameters $\lambda_{w_j v}$.

Repeating the above argument for each of the y_i, we obtain k linear equations in k unknowns. It remains to show that the system of equations is generically non-singular. This amounts to showing generic invertibility for the $k \times k$ matrix A with entries

$$A_{ij} = \Sigma_{y_i w_j} - \sum_{h \in H_i} \Sigma_{hw_j}\lambda_{hy_i}.$$

The invertibility of A follows from the existence of the half-trek system from Y to ω_{33} with no sided intersection and Lemma 4.3 below. We conclude that each $w_i \to v$ is generically (rationally) identifiable as claimed. □
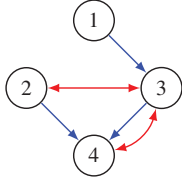
Figure 4: A graph that serves to illustrate differences between Theorem 1 of Chen [16] and our Theorem 4.1.

Remark 4.2. Our Theorem 4.1 generalizes Theorem 1 in [16] in two ways. Firstly, we make the trivial, but for our purposes important, modification to formulate our theorem in a fashion that is agnostic as to how prior generic identifications were obtained. For the presentation in [16] it was more natural to focus only on such identifications being obtained from prior applications of his theorem. Secondly, and more substantially, the results in [16] do not consider the possibility that, recalling the setting of Theorem 4.1, some of the edges incoming to v may be known to be generically identifiable; failing to use this information makes the conditions on the set Y more restrictive. Indeed, but for our first modification, our theorem reduces to the result in [16] if we replace condition (ii)(a) by the condition "π ends with an edge of the form $s \rightarrow v$ where $s \in W$."

As an example of how the above difference can appear in practice consider Figure 4 and suppose we have restricted the size of edge sets W we consider to be of size 1 (for larger graphs, this may be required for computational efficiency). Then, using $Y = \{1\}$ and $W = \{3\}$, one easily checks that $3 \rightarrow 4$ is generically identifiable. But now, showing that $2 \rightarrow 4$ is generically identifiable using $W = \{2\}$ is impossible using Theorem 1 in [16] because of the trek $2 \leftrightarrow 3 \rightarrow 4$ but this trek provides no problem for Theorem 4.1 as we have already shown that $3 \rightarrow 4$ is generically identifiable.

The following lemma generalizes Lemma 2 from [10] and completes the proof of Theorem 4.1.

Lemma 4.3. *Let $G = (V, D, B)$ be a mixed graph on n nodes with associated covariance matrix Σ. Moreover, let $S = \{s_1, \ldots, s_k\}$, $T = \{t_1, \ldots, t_k\} \subset V$. For every $1 \leq i \leq k$ let $H_i = \{h_1^i, \ldots, h_{\ell_i}^i\} \subset pa(s_i)$. Suppose there exists a half-trek system from S to T with no sided intersection. Then the $k \times k$ matrix A defined by*

$$A_{ij} = \Sigma_{s_i t_j} - \sum_{k=1}^{\ell_i} \Sigma_{h_k^i t_j} \lambda_{h_k^i s_i}$$

is generically invertible.

The proof of this lemma is deferred to Appendix B. Note that if let $W = pa(v)$ and strengthen condition (ii)(b) to require that all edges incoming to y be generically identifiable whenever there exists a half-trek from v to y, then Theorem 4.1 reduces to Theorem 2.10 of Foygel et al. [10], the usual half-trek identifiability theorem.

The conditions of Theorem 4.1 can be easily checked in polynomial time using max-flow computations, just as with the standard half-trek criterion. Unfortunately, in general, we do not know for which subset $W \subset pa(v)$ we should be checking the conditions of Theorem 4.1. This, in practice, means that we will have to check all subsets $W \subset pa(v)$. There are, of course, exponentially many such subsets in general. If we are in a setting where we may assume that all vertices have bounded in-degree, then checking all subsets requires only polynomial time. In the case that in-degrees are not bounded, we may also maintain polynomial time complexity by only considering subsets W of sufficiently large or small size. We provide pseudocode for an algorithm to iteratively identify the coefficients of a mixed graph leveraging Theorem 4.1 in Algorithm 1.

5 Edgewise generic nonidentifiability

In prior sections we have focused solely on sufficient conditions for demonstrating the generic identifiability of edges in a mixed graph. This, of course, begs the question of if there are any complementary necessary

---

Algorithm 1 Edgewise identification algorithm.

1: **Input:** A mixed graph $G = (V, D, B)$ with $V = \{1, \dots, n\}$ and a set of edges, *solved Edges*, known to be generically identifiable.
2: **repeat**
3:      **for** $v \leftarrow 1, \dots, n$ **do**
4:          $unsolved \leftarrow \{w \in V \mid w \rightarrow v \in G \text{ and } w \rightarrow v \notin solved\,Edges\}$.
5:          $maybeAllowed \leftarrow$
                $\{y \in V \backslash (\{v\} \cup \mathrm{sib}(v)) \mid z \in \mathrm{htr}(v) \cap \mathrm{pa}(y) \implies z \rightarrow y \in solved\,Edges\}$
6:          **for** $\varnothing \neq W \subset unsolved$ **do**
7:              $allowed \leftarrow \{y \in maybeAllowed \mid \mathrm{htr}(y) \cap \mathrm{tr}(\{p \in \mathrm{pa}(y) \mid p \rightarrow y \notin solved\,Edges\}) \cap unsolved \subset W\}$
8:              $exists \leftarrow$ Using max-flow computations, does there exist a half-trek system from *allowed* to W of size $|W|$ with no sided intersection?
9:              **if** *exists* is true **then**
10:                 $solved\,Edges \leftarrow solvedEdges \cup \{e \rightarrow v \mid e \in E\}$
11:                 Break out of the current loop
12:             **end if**
13:         **end for**
14:     **end for**
15: **until** No additional edges have been added to *solvedEdges* on the most recent loop.
16: **Output:** *solved Edges*, the set of edges found to be generically (rationally) identifiable.

---

conditions. That is, if there exist conditions that, when failed, show that a given edge is generically many-to-one. To our knowledge, the following is the only known necessary condition for generic identifiability and considers all parameters of a mixed graph G simultaneously.

Theorem 5.1 (Theorem 2 of Foygel et al. [10]). Suppose $G = (V, D, B)$ is a mixed graph in which every family $(Y_v : v \in V)$ of subsets of the vertex set V either contains a set Y_v that fails to satisfy the half-trek criterion with respect to v or contains a pair of sets (Y_v, Y_w) with $v \in Y_w$ and $w \in Y_v$. Then the parameterization ϕ_G is generically infinite-to-one.

This theorem operates by showing that, given its conditions, the Jacobian of the map ϕ_G fails to have full column rank and thus must have infinite-to-one fibers. Unfortunately this theorem does not give any indication regarding which edges are, in particular, generically infinite-to-one. The theorem below gives a simple condition which guarantees that a directed edge is generically infinite-to-one.

Theorem 5.2. Let $G = (V, D, B)$ be a mixed graph and let $v \rightarrow w \in G$. Suppose that for every $z \in V \backslash \{w\}$ we have either $z \leftrightarrow w \in G$ or v is not half-trek reachable from z. Let $\mathrm{proj}_{v \rightarrow w}$ be the projection $(\Lambda, \Omega) \mapsto \lambda_{vw}$ for $v \rightarrow w \in G$. Then $\mathrm{proj}_{v \rightarrow w}(\mathcal{F}(\Lambda, \Omega))$ is infinite for all $(\Lambda, \Omega) \in \Theta = \mathbb{R}^D_{reg} \times PD(B)$.

Proof. Let $(\Lambda, \Omega) \in \Theta$ and $\Sigma = \phi_G(\Lambda, \Omega) = (I - \Lambda)^{-T} \Omega (I - \Lambda)^{-1}$. We will show that for each matrix $\Gamma = (\gamma_{xy}) \in \mathbb{R}^D_{reg}$ that agrees with Λ in all but (possibly) the (v, w) entry, we can find $\Psi \in PD(B)$ for which $\phi_G(\Gamma, \Psi) = \Sigma$. The claim then follows by noting that the choices for Γ allow for infinitely many values of γ_{vw}.

Let $\Gamma \in \mathbb{R}^D_{reg}$ be as above, and let $x \neq y \in V$ be such that $x \leftrightarrow y \notin G$. Then

$$((I - \Gamma)^T \Sigma (I - \Gamma))_{xy}$$
$$= \sigma_{xy} - \sum_{z \in \mathrm{pa}(x)} \sigma_{yz} \gamma_{zx} - \sum_{z \in \mathrm{pa}(y)} \sigma_{xz} \gamma_{zy} + \sum_{z \in \mathrm{pa}(x)} \sum_{z' \in \mathrm{pa}(y)} \gamma_{zx} \gamma_{z'y} \sigma_{zz'}.$$

Whenever $x, y \neq w$ then $\gamma_{zx} = \lambda_{zx}$ and $\gamma_{zy} = \lambda_{zy}$ in the above equation. Thus

$$0 = \Omega_{xy} = ((I - \Lambda)^T \Sigma (I - \Lambda))_{xy} = ((I - \Gamma)^T \Sigma (I - \Gamma))_{xy}.$$
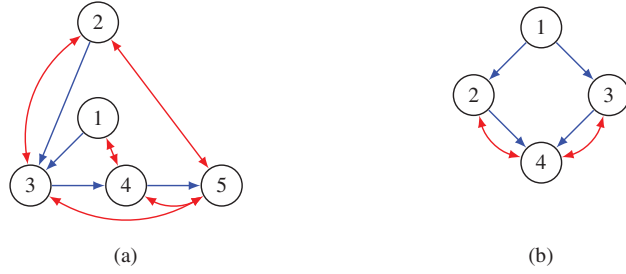
Figure 5: Two graphs serving to illustrate Theorem 5.2. (a) A graph in which all directed edges are identifiable except $2 \to 3$. The $2 \to 3$ edge can be shown to be infinite-to-one using Theorem 5.2. (b) A graph known to have generically infinite-to-one parameterization by Theorem 5.1 but for which Theorem 5.2 applies to no edge.

Next suppose, without loss of generality, that $x = w$ and $y \neq w$. Then, since y is a non-sibling of w, we must have that v is not half-trek reachable from y, and hence $\sigma_{vy} = \sum_{z \in \text{pa}(y)} \sigma_{vz} \lambda_{zy}$. But then

$$((I - \Gamma)^T \Sigma (I - \Gamma))_{wy}$$
$$= \sigma_{wy} - \sum_{z \in \text{pa}(w)} \sigma_{yz} \gamma_{zw} - \sum_{z \in \text{pa}(y)} \sigma_{wz} \gamma_{zy} + \sum_{z \in \text{pa}(w)} \sum_{z' \in \text{pa}(y)} \gamma_{zw} \gamma_{z'y} \sigma_{zz'}$$
$$= -\sigma_{vy} \gamma_{vw} + \sum_{z' \in \text{pa}(y)} \gamma_{vw} \lambda_{z'y} \sigma_{vz'}$$
$$+ \sigma_{wy} - \sum_{v \neq z \in \text{pa}(w)} \sigma_{yz} \lambda_{zw} - \sum_{z \in \text{pa}(y)} \sigma_{wz} \lambda_{zy} + \sum_{v \neq z \in \text{pa}(w)} \sum_{z' \in \text{pa}(y)} \lambda_{zw} \lambda_{z'y} \sigma_{zz'}$$
$$= -\gamma_{vw} \left(\sigma_{vy} - \sum_{z' \in \text{pa}(y)} \lambda_{z'y} \sigma_{vz'} \right)$$
$$+ \sigma_{wy} - \sum_{v \neq z \in \text{pa}(w)} \sigma_{yz} \lambda_{zw} - \sum_{z \in \text{pa}(y)} \sigma_{wz} \lambda_{zy} + \sum_{v \neq z \in \text{pa}(w)} \sum_{z' \in \text{pa}(y)} \lambda_{zw} \lambda_{z'y} \sigma_{zz'}.$$

Now since $\sigma_{vy} - \sum_{z' \in \text{pa}(y)} \lambda_{z'y} \sigma_{vz'} = 0$, we have that $0 = -\gamma_{vw} (\sigma_{vy} - \sum_{z' \in \text{pa}(y)} \lambda_{z'y} \sigma_{vz'}) = -\lambda_{vw} (\sigma_{vy} - \sum_{z' \in \text{pa}(y)} \gamma_{z'y} \sigma_{vz'})$. Therefore,

$$((I - \Gamma)^T \Sigma (I - \Gamma))_{wy}$$
$$= -\lambda_{vw} \left(\sigma_{vy} - \sum_{z' \in \text{pa}(y)} \lambda_{z'y} \sigma_{vz'} \right)$$
$$+ \sigma_{wy} - \sum_{v \neq z \in \text{pa}(w)} \sigma_{yz} \lambda_{zw} - \sum_{z \in \text{pa}(y)} \sigma_{wz} \lambda_{zy} + \sum_{v \neq z \in \text{pa}(w)} \sum_{z' \in \text{pa}(y)} \lambda_{zw} \lambda_{z'y} \sigma_{zz'}$$
$$= ((I - \Lambda)^T \Sigma (I - \Lambda))_{wy}$$
$$= \Omega_{wy} = 0.$$

Let $\Psi = (I - \Gamma)^T \Sigma (I - \Gamma)$. We have just shown that $\Psi_{xy} = 0$ for every $x, y \in V$ such that $x \leftrightarrow y \notin G$. To see that $\Psi \in PD(B)$ it remains to show that Ψ is positive definite. But this is obvious from its definition since Σ is positive definite and $I - \Gamma$ is invertible. We conclude that $\phi_G(\Gamma, \Psi) = \Sigma$ which proves the claim. □

Example 5.3. Let G be the graph in Figure 5a. Using the necessary condition of the HTC, Theorem 5.1, we find that ϕ_G is generically infinite-to-one. To identify which edges of G are themselves infinite-to-one we use Theorem 5.2. Doing so, one easily finds that the $2 \to 3$ edge of G is generically infinite-to-one. Indeed, using the edgewise identification techniques of Section 4, we see that all other directed edges of G are generically identifiable so we have completely characterized which directed edges of G are, and are not, generically identifiable.

We stress, however, that Theorem 5.2 does not imply Theorem 5.1; that is, there are graphs G for which Theorem 5.1 shows ϕ_G is infinite-to-one but Theorem 5.2 cannot verify that any edges of G are infinite-to-one. For example, see Figure 5b.

6 Computational experiments

In this section we will provide some computational experiments that demonstrate the usefulness of our theorems in extending the applicability of the half-trek criterion. All of our following experiments are carried out in the R programming language and the following algorithms are implemented in our R package SEMID which is available on CRAN, the Comprehensive R Archive Network [22, 23], as well as on GitHub.[1] We will be considering four different identification algorithms for checking generic identifiability:

  (i)   The standard half-trek criterion (HTC) algorithm.
 (ii)   The edgewise identification (EID) algorithm, displayed in Algorithm 1, where the input set of *solved Edges* is empty.
(iii)   The trek-separation identification (TSID) algorithm. Similarly as for Algorithm 1 this algorithm iteratively applies Theorem 3.8 until it fails to identify any additional edges. (Since we are considering a small number of nodes there is no need to limit the size of sets S and T we are searching for in our computation.)
(iv)   The EID+TSID algorithm. This algorithm alternates between the EID and TSID algorithms until it fails to identify any additional edges.

We emphasize that when all of the directed edges, i.e., the matrix Λ is generically (rationally) identifiable then we also have that $\Omega = (I - \Lambda)^T \Sigma (I - \Lambda)$ is generically (rationally) identifiable.

In Table 1 from [24], the authors list all 112 acyclic non-isomorphic mixed graphs on 5 nodes which are generically identifiable but for which the half-trek criterion remains inconclusive even when using decomposition techniques. We run the EID, TSID, and EID+TSID algorithms upon the 112 inconclusive graphs and find that 23 can be declared generically identifiable by the EID algorithm, 0 by the TSID algorithm, and 98 by the EID+TSID algorithm. Thus it is only by using both the determinantal equations discovered by t-separation and the edgewise identification techniques that one sees a substantial increase in the number of graphs that can be declared generically identifiable.

We observe a similar trend to the above when allowing cyclic mixed graphs. In Table 2 of [24], the authors list 75 randomly chosen, cyclic (i.e., containing a loop in the directed part), mixed graphs that are known to be rationally identifiable but cannot be certified so by the half-trek criterion. Of these 75 graphs, 4 are certified to be generically identifiable by the EID algorithm, 0 by the TSID algorithm, and 34 by the EID+TSID algorithm.

A listing of the 14 acyclic and 41 cyclic mixed graphs that could not be identified by the EID+TSID algorithm are listed as integer pairs $(d, b) \in \mathbb{N}^2$ in Table 1. The algorithm to convert a pair (d, b) in that table to a mixed graph G on n nodes is

1. For $v \leftarrow 1, \ldots, n$, for $w \leftarrow 1, \ldots, v - 1, v + 1, \ldots, n$, do
       Add edge $v \rightarrow w$ to G if $d \bmod 2 = 1$
       Replace d with $\lfloor d/2 \rfloor$
2. For $v \leftarrow 1, \ldots, n - 1$, for $w \leftarrow v + 1, \ldots, n$, do
       Add edge $v \leftrightarrow w$ to G if $b \bmod 2 = 1$
       Replace b with $\lfloor b/2 \rfloor$

See Figure 6 for an example of a cyclic and acyclic graph that the EID+TSID algorithm fails to correctly certify as generically identifiable.

---

1 See https://github.com/Lucaweihs/SEMID.

Table 1: Of the 112 acyclic and 75 cyclic mixed graphs on 5 nodes described in Tables 1 and 2 from [24], we display the 12 acyclic and 41 cyclic graphs which are known to be generically identifiable but for which the EID+TSID algorithm could not certify that all edges were generically identifiable. Each graph is encoded as a pair (d, b), see text for details.

Acyclic		Cyclic	
(4456, 113)	(345, 440)	(6629, 512)	(75321, 516)
(360, 117)	(71329, 18)	(74536, 788)	(75398, 20)
(6275, 172)	(81089, 0)	(5545, 96)	(70803, 896)
(6307, 172)	(4714, 41)	(75112, 72)	(4457, 592)
(6275, 188)	(70881, 80)	(74970, 4)	(74883, 522)
(360, 369)	(74963, 512)	(4579, 384)	(350, 112)
(4696, 401)	(74886, 268)	(70594, 65)	(74883, 2)
(4936, 401)	(5058, 304)	(74921, 66)	(74950, 260)
(4936, 402)	(70821, 513)	(70474, 640)	(74890, 38)
(4680, 403)	(74915, 6)	(74922, 66)	(81076, 0)
(840, 466)	(5267, 82)	(13160, 65)	(70851, 32)
(5257, 658)	(76852, 128)	(4938, 448)	(1430, 120)
(5257, 659)	(71075, 516)	(4730, 640)	(5251, 418)
(4680, 914)	(4397, 897)	(70358, 1)	

(a)   (b)
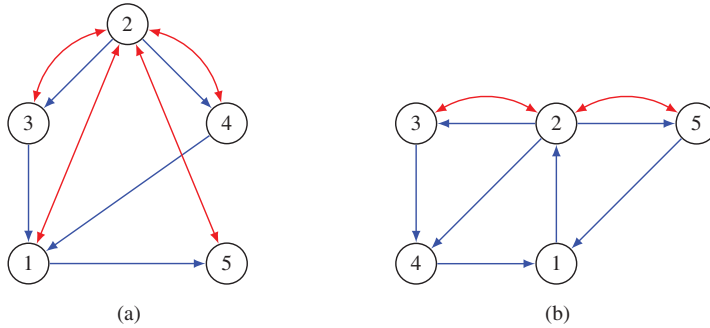
Figure 6: Two graphs for which the EID+TSID algorithm is inconclusive. (a) is acyclic while (b) contains a cycle.

7 Conclusion

By exploiting the trek-separation characterization of the vanishing of subdeterminants of the covariance matrix Σ corresponding to a mixed graph G, we have shown that individual edge coefficients can be generically identified by quotients of subdeterminants. This constitutes a generalization of instrumental variable techniques that are derived from conditional independence. We have also shown how this information, in concert with a generalized half-trek criterion, allows us to prove that substantially more graphs have all or some subset of their parameters generically identifiable.

Our work on identification by ratios of determinants focuses on a single edge coefficient. However, it seems possible to give a generalization that is in the spirit of the generalized instrumental sets from [15]; see also [25]. These leverage several conditional independencies to find a linear equation system that can be used to identify several edge coefficients simultaneously, under specific assumptions on the interplay of the conditional independencies and the edges to be identified. We illustrate the idea of how to do this using general determinants in the following example. However, a full exploration of this idea is beyond the scope of this paper. In particular, we are still lacking mathematical tools that, in suitable generality, could be used to certify that constructed linear equation systems have a unique solution.

Example 7.1. Let G be the graph in Figure 7 with corresponding covariance matrix $\Sigma = (I - \Lambda)^{-T}\Omega(I - \Lambda)^{-1}$. Then, by similar considerations to those in Example 3.4, one may show that
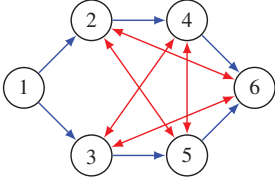
Figure 7: A graph where the edges $4 \to 6$ and $5 \to 6$ can be simultaneously proven to be generically identifiable by solving a 2×2 linear system of determinantal equations.

$$\begin{pmatrix} |\Sigma_{\{3,5\},\{1,4\}}| & |\Sigma_{\{3,5\},\{1,5\}}| \\ |\Sigma_{\{2,4\},\{1,4\}}| & |\Sigma_{\{2,4\},\{1,5\}}| \end{pmatrix} \begin{pmatrix} \lambda_{46} \\ \lambda_{56} \end{pmatrix} = \begin{pmatrix} |\Sigma_{\{3,5\},\{1,6\}}| \\ |\Sigma_{\{2,4\},\{1,6\}}| \end{pmatrix}.$$

Using computer algebra we find that the 2×2 matrix on the left hand side of the above equation has all non-zero polynomial entries, so that this is not equivalent to simply applying Theorem 3.8 for $4 \to 6$ and $5 \to 6$ separately, and has non-zero determinant. It follows that the above system is generically invertible and thus λ_{46} and λ_{56} are generically identifiable.

Appendix

A Proof of Lemma 3.7

We will require a known generalization of the Gessel-Viennot-Lindström lemma which we now state.

Definition A.1. Let $G = (V, D)$ be a directed graph with vertices $V = \{1, \ldots, n\}$ and corresponding matrix of indeterminants Λ. Let $\pi = v_1 \to v_2 \to \cdots \to v_\ell$ be a directed path in G. Then define the *loop erased* path $LE(\pi)$ corresponding to π recursively as follows. If π contains no loops then $\pi = LE(\pi)$. Otherwise there exist indices $1 \le i < j \le \ell$ such that $v_i = v_j$. Then $LE(\pi) = LE(\pi')$ where $\pi' = v_1 \to v_2 \to \cdots \to v_i \to v_{j+1} \to \cdots \to v_\ell$. It can be shown that $LE(\pi)$ is well defined (i.e. is independent of the ordering of the above recursion).

Lemma A.2 (Gessel-Viennot-Lindström Generalization, Theorem 6.1 from [26]). *Let $G = (V, D)$ be a directed graph with vertices $V = \{1, \ldots, n\}$ and corresponding matrix of indeterminants Λ. Define $\Psi = (I - \Lambda)^{-1}$ and for any directed path π in G define the path polynomial $\pi(\Lambda) = \prod_{w \to v \in \pi} \lambda_{wv}$. Then for any $S = \{s_1, \ldots, s_k\}$, $T = \{t_1, \ldots, t_k\} \subset V$ we have that*

$$|\Psi_{S,T}| = \sum_{\tau \in P_n} sign(\tau) \sum_{\substack{s_1 \xrightarrow{\pi_1} t_{\tau(1)}, \ldots, s_k \xrightarrow{\pi_k} t_{\tau(k)} \\ i < j \implies \pi_j \cap LE(\pi_i) = \varnothing}} \pi_1(\Lambda) \cdots \pi_k(\Lambda),$$

here the above inner sum is over all directed path systems $\Pi = \{\pi_1, \ldots, \pi_k\}$ with π_i going from s_i to $t_{\tau(i)}$ for all i, where π_j and $LE(\pi_i)$ share no vertices for $i < j$. Hence $|\Psi_{S,T}| = 0$ if and only if every system of directed paths from S to T has two paths which share a vertex.

The remaining proof of Lemma 3.7 proceeds in several parts and closely follows similar results in [14] and [19]. As such we will state several lemmas whose proofs require only small modifications of existing results

(such as replacing the standard Gessel-Viennot-Lindström Lemma with its generalization above). In such cases we will simply direct the reader to the corresponding proof and sketch the necessary modifications.

Definition A.3. Let $G = (V, D, B)$ be a mixed graph and let $U \subset D$. We say a trek π in G *avoids U on the left (right)* if the left (right) side of π uses no edges from U. Similarly we say a system of treks Π in G *avoids U on the left (right)* if every trek $\pi \in \Pi$ avoids U on the left (right). If $U_L, U_R \subset D$ we say that a *trek (or trek system) avoids (U_L, U_R)* if it avoids U_L on the left and U_R on the right.

Lemma A.4. *Let $G = (V, D, B)$ be a mixed graph and let Λ, Ω be $n \times n$ matrices of indeterminants corresponding to the directed and bidirected parts of G respectively. Suppose that $B = \varnothing$ so that Ω is diagonal. Letting $D_L, D_R, \Lambda^L, \Lambda^R, \Gamma$, and G^*_{flow} be as in Lemma 3.7 we have that for any $S, T \subset V$ with $|S| = |T| = k$, $|\Gamma_{S,T}| = 0$ if and only if the max-flow from S to T' in G^*_{flow} is $< k$.*

Proof. In the following, whenever we say "As in x," we mean "As in the proof of x in Sullivant et al. [14]."

As in Lemma 3.2, we have $|\Gamma_{S,T}| = 0$ if and only if for every set $A \subset V$ with $|A| = K$ we have $|((I - \Lambda^L)^{-1})_{S,A}| = 0$ or $|((I - \Lambda^R)^{-1})_{A,T}| = 0$. As in Prop. 3.5, using the above result, and applying our version of the Gessel-Viennot-Lindström Lemma, we have that $|\Gamma_{S,T}| = 0$ if and only if every system of (simple) treks avoiding $(D \backslash D_L, D \backslash D_R)$ has sided intersection.

Now noticing that $B = \varnothing$ simplifies the definition of G^*_{flow}, we have as in Prop. 3.5 that the (simple) treks from u to v avoiding $(D \backslash D_L, D \backslash D_R)$ in G are in bijective correspondence with directed paths from u to v' in G^*_{flow}. Finally the result follows by noticing that max-flow systems from S to T' in G^*_{flow} of size k correspond to systems of treks from S to T avoiding $(D \backslash D_L, D \backslash D_R)$ with no-sided intersection (that is, if one exists so does the other). Combining the above if and only if statements, the result then follows. □

We have now proven our desired result in the case $B = \varnothing$, it remains to show that this implies the case $B \neq \varnothing$. To this end, we say that $\widetilde{G} = (\widetilde{V}, \widetilde{D}, \widetilde{B})$ is the *bidirected subdivision* of $G = (V, D, B)$ if it equals G but where we have replaced every bidirected edge $i \leftrightarrow j \in G$ with a vertex $v_{(i,j)}$ and two edges $v_{(i,j)} \to i$ and $v_{(i,j)} \to j$ (with associated parameters $\widetilde{\omega}_{(i,j),(i,j)}, \widetilde{\lambda}_{(i,j)i}, \widetilde{\lambda}_{(i,j)j}$). Note that we have subdivided every bidirected edge into two directed edges which motivates the naming convention. Let $\widetilde{D}_L$ and $\widetilde{D}_R$ be equal to D_L and D_R respectively but where we have also added in the new edges $v_{(i,j)} \to i$ and $v_{(i,j)} \to j$ for every $i \leftrightarrow j \in G$. Let $\widetilde{\Lambda}, \widetilde{\Omega}$ be matrices of indeterminants corresponding to $\widetilde{G}$ and let $\widetilde{\Lambda}^L, \widetilde{\Lambda}^R$ correspond to $\widetilde{D}_L, \widetilde{D}_R$ just as for G. We now have the following result that relates G and $\widetilde{G}$.

Lemma A.5. *Let $G, \widetilde{G}$ be as in the prior paragraph. Then letting $\widetilde{\Gamma} = (I - \widetilde{\Lambda}^L)^{-T} \widetilde{\Omega} (I - \widetilde{\Lambda}^R)^{-1}$ we have that, for any polynomial f taking, as input, an $n \times n$ matrix of variables, we have that $f(\Gamma) = 0$ if and only if $f(\widetilde{\Gamma}) = 0$. In particular, since the subdeterminant of a matrix is a polynomial in the entries of the matrix, we have that for any $S, T \subset V$ with $|S| = |T| = k$, $|\Gamma_{S,T}| = 0$ if and only if $|\widetilde{\Gamma}_{S,T}| = 0$.*

Proof. This proof follows, essentially exactly, as the first part of the proof of Prop. 2.5 in Draisma et al. [19]. □

Now we show that the above subdivision trick produces a graph $\widetilde{G}^*_{flow}$ for which the max-flow between vertex sets is the same as for G^*_{flow}.

Lemma A.6. *Consider the graphs $G^*_{flow} = (V^*, D^*)$ from the Lemma 3.7 statement and let $\widetilde{G}^*_{flow} = (\widetilde{V}^*, \widetilde{D}^*)$ be corresponding flow graph for the bidirected subdivision $\widetilde{G}$ of G. Let $S = \{s_1, \ldots, s_k\}$, $T = \{t_1, \ldots, t_k\} \subset V$. Then the maximum flow from S to $T' = \{t'_1, \ldots, t'_k\}$ in G^*_{flow} equals the maximum flow from S to T' in $\widetilde{G}^*_{flow}$.*

Proof. Recall that a flow system on a graph is an assignment of flow to the edges and vertices of the graph satisfying the usual flow constraints. Also recall that, for graphs with integral capacities, there always exists a max-flow system between subsets of nodes for which all flow assignments upon edges and vertices take values in $\mathbb{N}$. We will show that any (integral valued) max-flow system from S to T' in $\widetilde{G}^*_{flow}$ corresponds to a unique flow system in G^*_{flow} with the same total flow and vice-versa. Our result then follows.

Let $\widetilde{\mathscr{F}}$ be a max-flow system from S to T' on $\widetilde{G}^*_{\text{flow}}$ from S to T' with integral flow assignments. Since $\widetilde{G}^*_{\text{flow}}$ and G^*_{flow} have all capacities equal to 1 it follows that $\widetilde{\mathscr{F}}$ assigns either 0 or 1 flow to all edges and vertices in the graph.

We now construct a flow system $\mathscr{F}$ on G^*_{flow} with the same capacity. First let $\mathscr{F}$ assign the same capacity to all edges and vertices that $\mathscr{F}$ shares with $\widetilde{\mathscr{F}}$. Note that if $\widetilde{\mathscr{F}}$ does not assign any flow to any of the edges incoming to the vertices $v_{(i,j)}$ then $\mathscr{F}$ already corresponds to a flow system on G^*_{flow} with the same total flow. Suppose otherwise that $\widetilde{\mathscr{F}}$ assigns 1 unit of flow to the edges $\{a_1 \to v_{a_1 b'_1}, \ldots, a_k \to v_{a_k b'_k}\}$. Since $v_{(i,j)}$ and the a_i have capacity 1 it follows that $a_i \neq a_j$ and $v_{a_i b'_i} \neq v_{a_i b'_i}$ for all $i \neq j$. For each edge $a_i \to v_{a_i b'_i}$, since $v_{a_i b'_i}$ has two outgoing edges $v_{a_i b'_i} \to a'_i$ and $v_{a_i b'_i} \to b'_i$, there are two possible cases:

– Case 1: $\widetilde{\mathscr{F}}$ assigns 1 flow to $v_{a_i b'_i} \to a'_i$.
   In this case assign a flow of 1 to the edge $a_i \to a'_i$ in $\mathscr{F}$.
– Case 2: $\widetilde{\mathscr{F}}$ assigns 1 flow to $v_{a_i b'_i} \to b'_i$.
   In this case assign a flow of 1 to the edge $a_i \to b'_i$ in $\mathscr{F}$.

It is easy to check that $\mathscr{F}$ is indeed a valid flow system on G^*_{flow} with the same flow as $\widetilde{\mathscr{F}}$.

To see the oppose direction let $\mathscr{F}$ be a max-flow system from S to T' on G^*_{flow} from S to T' with integral flow assignments. We now construct a flow system $\widetilde{\mathscr{F}}$ on $\widetilde{G}^*_{\text{flow}}$ with the same capacity. As before, first let $\widetilde{\mathscr{F}}$ assign the same capacity to all edges and vertices that $\widetilde{\mathscr{F}}$ shares with $\mathscr{F}$. Note that if $\mathscr{F}$ does not assign any flow to any of the edges $a \to b'$ for $(a, b) \in B$ then $\widetilde{\mathscr{F}}$ already corresponds to a flow system on G^*_{flow} with the same total flow. Suppose otherwise that $\widetilde{\mathscr{F}}$ assigns 1 unit of flow to the edges $E = \{a_1 \to b'_1, \ldots, a_k \to b'_k\}$ with $(a_i, b_i) \in B$ for all i. Since all vertices in $\mathscr{F}$ have capacity 1 we must have that $a_i \neq a_j$ and $b_i \neq b_j$ for all $i \neq j$. There are two possible cases:

– Case 1: $a_i \to b'_i \in E$ and $b_i \to a_i \notin E$.
   In this case assign a flow of 1 along the path $a_i \to v_{a_i b_i} \to b'_i$ in $\widetilde{\mathscr{F}}$.
– Case 2: $a_i \to b'_i \in E$ and $b_i \to a_i \in E$.
   In this case assign a flow of 1 to the edges $a_i \to a'_i$ and $b_i \to b'_i$ in $\widetilde{\mathscr{F}}$.

One may now check that $\widetilde{\mathscr{F}}$ is a valid flow system on $\widetilde{G}^*_{\text{flow}}$ with the same flow as $\mathscr{F}$. □

Finally we are in a position to easily prove Lemma 3.7. Note that, by Lemma A.5 we have that $|\Gamma_{S,T}| = 0$ if and only if $|\widetilde{\Gamma}_{S,T}| = 0$. By Lemma A.4 we have that $|\widetilde{\Gamma}_{S,T}| = 0$ if and only if the max-flow from S to T' in $\widetilde{G}^*_{\text{flow}}$ equals $|S| = k$. Finally Lemma A.6 gives us that the max-flow from S to T' in $\widetilde{G}^*_{\text{flow}}$ equals the max-flow from S to T' in G^*_{flow}. Hence we have that $|\Gamma_{S,T}| = 0$ if and only if the max-flow from S to T' in G^*_{flow} equals k, this was our desired statement.

B Proof of Lemma 4.3

The proof of this lemma follows almost identically as the proof of Lemma 2 in [10]. We simply restate the arguments there in our setting. For any $v, w \in V$ let $\mathscr{H}(v, w)$ be the set of half treks from v to w in G. Also let $\mathscr{T}_{ij}$ be the set of all treks from s_i to t_j in G which do not begin with an edge of the form $s_i \leftarrow h^i_k$ for any $1 \leq k \leq \ell_i$. Then it is easy to see that $\mathscr{H}(s_i, t_j) \subset \mathscr{T}_{ij}$. Now, by the Trek Rule (Proposition 2.5), we have that

$$A_{ij} = \sum_{\pi \in \mathscr{T}_{ij}} \pi(\Lambda, \Omega).$$

Now for any system of treks Π define the monomial

$$\Pi(\Lambda, \Omega) = \prod_{\pi \in \Pi} \pi(\Lambda, \Omega).$$

Then, by Leibniz's formula for the determinant, we have that

$$|A| = \sum_{\Pi} (-1)^{\text{sign}(\Pi)} \Pi(\Lambda, \Omega) \tag{15}$$

where the above sum is over all trek systems Π from S to T using treks only in the set $\cup_{1 \le i,j \le k} \mathscr{T}_{ij}$; here the sign($\Pi$) is the sign of the permutation that writes $t_1, \ldots, t_k$ in the order of their appearance as targets of the treks in Π.

By assumption, there exists a half-trek system from S to T with no-sided intersection. Since such a system exists, let Π be a half-trek system of minimum total length among all such half-trek systems. Since $\mathscr{H}(s_i, t_j) \subset \mathscr{T}_{ij}$ for all i, j it follows that Π is included as one of the trek systems in the summation (15). Let Ψ be any system of treks from S to T such that $\Psi(\Lambda, \Omega) = \Pi(\Lambda, \Omega)$. Lemma 1 from [10] proves that we must have $\Psi = \Pi$ so that Π is the unique system of treks from S to T with corresponding trek monomial $\Pi(\Lambda, \Omega)$. It thus follows that the coefficient of the monomial $\Pi(\Lambda, \Omega)$ in $|A|$ is $(-1)^{\text{sign}(\Pi)}$ and thus $|A|$ is not the zero polynomial (or power series if the sum is infinite). Hence, for generic choices of (Λ, Ω), we have that $|A| \ne 0$ so that A is generically invertible.

References

1. Bollen, KA. Structural equations with latent variables, Wiley Series in Probability and Mathematical Statistics: Applied Probability and Statistics. John Wiley & Sons, Inc., New York, a Wiley-Interscience Publication, 1989.
2. Pearl J. Causality models, reasoning, and inference, 2nd edn. Cambridge: Cambridge University Press, 2009.
3. Spirtes P, Glymour C, Scheines R. Causation, prediction, and search, 2nd edn. Cambridge, MA: MIT press, 2000.
4. Wright S. Correlation and causation. J. Agricultural Res. 921;20:557–585.
5. Wright S. The method of path coefficients. Ann. Math. Statist. 1934;5:161–215.
6. Didelez V, Meng S, Sheehan NA. Assumptions of IV methods for observational epidemiology. Statist. Sci. 2010;25:22–40.
7. Drton M, Foygel R, Sullivant S. Global identifiability of linear structural equation models. Ann. Statist. 2011;39:865–886.
8. Shpitser I, Pearl J. Identification of joint interventional distributions in recursive semi-Markovian causal models. In: Proceedings of the 21st National Conference on Artificial Intelligence. Menlo Park, CA: AAAI Press, 2006:1219–1226.
9. Tian J, Pearl J. A general identification condition for causal effects. In: Proceedins of the 18th national conference on artificial intelligence. Menlo Park, CA: AAAI Press, 2002:567–573.
10. Foygel R, Draisma J, Drton M. Half-trek criterion for generic identifiability of linear structural equation models. Ann. Statist. 2012a;40:1682–1713.
11. Chen, B. Decomposition and identification of linear structural equation models. ArXiv e-prints, 1508.01834, 2015.
12. Chen, B, Tian J, Pearl J. Testable implications of linear structural equations models. In: Brodley CE, Stone P, editors. Proceedings of the twenty-eighth AAAI conference on artificial intelligence. AAAI Press, 2014:2424–2430.
13. Drton M, Weihs L. Generic identifiability of linear structural equation models by ancestor decomposition. Scandinavian J Stat. 2016;43:1035–1045.
14. Sullivant S, Talaska K, Draisma J. Trek separation for Gaussian graphical models. Ann. Statist. 2010;38:1665–1685.
15. Brito, C, Pearl J. Generalized instrumental variables. In: Proceedings of the eighteenth conference annual conference on uncertainty in artificial intelligence (UAI-02). San Francisco, CA: Morgan Kaufmann, 2002:85–93.
16. Chen, B. Identification and overidentification of linear structural equation models. In: Lee DD, Sugiyama M, Luxburg UV, Guyon I, Garnett R, editors. Advances in neural information processing systems 29. Curran Associates, Inc., 2016: 1579–1587.
17. Drton M. Algebraic problems in structural equation modeling. ArXiv:1612.05994, 2016.
18. Okamoto M. Distinctness of the eigenvalues of a quadratic form in a multivariate sample. Ann. Statist. 1973;1:763–765.
19. Draisma, J, Sullivant S, Talaska K. Positivity for Gaussian graphical models. Adv. in Appl. Math. 2013;50:661–674.
20. Cormen, TH, Leiserson CE, Rivest RL, Stein C. Introduction to algorithms, 3rd ed. MIT Press, Cambridge, MA, 2009.
21. van der Zander B, Textor J, Liśkiewicz M. Efficiently finding conditional instruments for causal inference. In: Proceedings of the 24th international joint conference on artificial intelligence (IJCAI 2015). AAAI Press, 2015:3243–3249.
22. Foygel R, Drton M. SEMID: Identifiability of linear structural equation models, r package version 0.1, 2013.
23. R Core Team. R: A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing, 2014.

24. Foygel R, Draisma J, Drton M. Supplement to half-trek criterion for generic identifiability of linear structural equation models. Ann. Statist. 2012b;40.
25. van der Zander B, Liśkiewicz M. On searching for generalized instrumental variables. In: Proceedings of the 19th international conference on artificial intelligence and statistics (AISTATS 2016), JMLR Proceedings, 2016:1214–1222.
26. Fomin S. Loop-erased walks and total positivity. Trans. Amer. Math. Soc. 2001;353:3563–3583 (electronic).