Title: Towards standard imsets for maximal ancestral graphs

Speakers: Robin Evans

Collection: Causal Inference & Quantum Foundations Workshop

Date: April 17, 2023 - 12:00 PM

URL: https://pirsa.org/23040105

Abstract: "Imsets, introduced by Studený (see Studený, 2005 for details), are an algebraic method for representing conditional independence models. They have many attractive properties when applied to such models, and they are particularly nice when applied to directed acyclic graph (DAG) models. In particular, the standard imset for a DAG is in one-to-one correspondence with the independence model it induces, and hence is a label for its Markov equivalence class. We present a proposed extension to standard imsets for maximal ancestral graph (MAG) models, which have directed and bidirected edges, using the parameterizing set representation of Hu and Evans (2020). By construction, our imset also represents the Markov equivalence class of the MAG.

We show that for many such graphs our proposed imset defines the model, though there is a subclass of graphs for which the representation does not. We prove that it does work for MAGs that include models with no adjacent bidirected edges, as well as for a large class of purely bidirected models. If there is time, we will also discuss applications of imsets to structure learning in MAGs.

This is joint work with Zhongyi Hu (Oxford).

References

Z. Hu and R.J. Evans, Faster algorithms for Markov equivalence, In Proceedings for the 36th Conference on Uncertainty in Artificial Intelligence (UAI-2020), 2020.

M. Studený, Probabilistic Conditional Independence Structures, Springer-Verlag, 2005."

Towards Standard Imsets for Maximal Ancestral Graphs

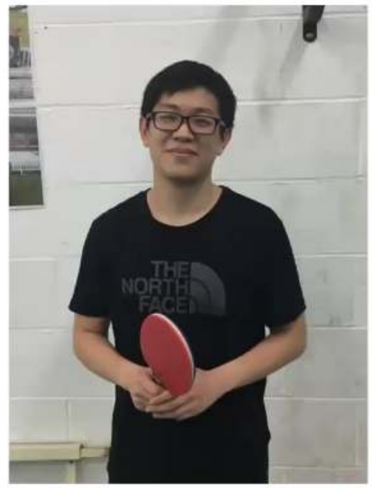
Robin Evans, University of Oxford

Causal Inference and Quantum Foundations
Perimeter Institute
17th April 2023

UNIVERSITY OF OXFORD

DEPARTMENT OF STATISTICS

Collaborators



Zhongyi Hu, University of Oxford

Outline

1. Imsets
2. DAG Models
3. MAG Models
4. Model Search

Outline

1. Imsets
2. DAG Models
3. MAG Models
4. Model Search

Imsets

Imsets were introduced by Studený (1995), as a method for representing arbitrary conditional independence models.

Let $\mathcal{P}(V)$ be the **power-set** of a finite set V.

Definition

An **imset** is an **i**nteger-valued **m**ulti**set**, or in other words a function

$$u : \mathcal{P}(V) \to \mathbb{Z}.$$

Imsets

Imsets were introduced by Studený (1995), as a method for representing arbitrary conditional independence models.

Let $\mathcal{P}(V)$ be the **power-set** of a finite set V.

Definition

An **imset** is an **i**nteger-valued **m**ulti**set**, or in other words a function

$$u : \mathcal{P}(V) \to \mathbb{Z}.$$

Since they are often sparse, we tend to represent them with combinations of identity functions:

$$\delta_A(X) = \begin{cases} 1 & \text{if } X = A, \\ 0 & \text{otherwise.} \end{cases}$$

Conditional Independence Models

Definition

We identify a **semi-elementary imset** with a triple (A, B, C) where

$$u_{\langle A,B|C\rangle} = \delta_C - \delta_{A\cup C} - \delta_{B\cup C} + \delta_{A\cup B\cup C}.$$

$u_{\langle A,B|C\rangle}$ represents the conditional independence $X_A \perp\!\!\!\perp X_B \mid X_C$.

Notice this conditional independence is equivalent to:

$$p(x_{ABC}) \cdot p(x_C) = p(x_{AC}) \cdot p(x_{BC})$$
$$\log p(x_C) - \log p(x_{AC}) - \log p(x_{BC}) + \log p(x_{ABC}) = 0.$$

Now we can see the analogy to the log-factorization.

Indeed, one can **test** a conditional independence by using the **interaction information operator** $\mathrm{I}_p : \mathcal{P}(V) \to \mathbb{R}$, and we have that $X_A \perp\!\!\!\perp X_B \mid X_C$ if and only if

$$\langle \mathrm{I}_P, u_{\langle A,B|C\rangle}\rangle = \mathrm{I}(p(x_C)) - \mathrm{I}(p(x_{AC})) - \mathrm{I}(p(x_{BC})) + \mathrm{I}(p(x_{ABC})) = 0.$$

Structural Imsets

Definition

An imset u is said to be **structural** if there exists some natural number k such that we can write

$$k \cdot u = \sum_{v \in \mathcal{I}(V)} k_v \cdot v, \qquad k_v \in \mathbb{N} \cup \{0\},$$

where $\mathcal{I}(V)$ is the collection of (semi-)elementary imsets over the variables in the set V.

Structural imsets can be said to represent a model.

Models

Definition

Given an independence $X_A \perp\!\!\!\perp X_B \mid X_C$, we say that it is **represented in a structural imset** u over V (and write $A \perp\!\!\!\perp B \mid C\,[u]$) if there exists $k \in \mathbb{N}$ such that

$$k \cdot u - u_{\langle A,B|C \rangle}$$

is also structural.

Can be tested with an integer linear program (Bouckaert et al., 2010).

Models

Definition

Given an independence $X_A \perp\!\!\!\perp X_B \mid X_C$, we say that it is **represented in a structural imset** u over V (and write $A \perp\!\!\!\perp B \mid C\,[u]$) if there exists $k \in \mathbb{N}$ such that

$$k \cdot u - u_{\langle A,B|C\rangle}$$

is also structural.

Can be tested with an integer linear program (Bouckaert et al., 2010).

Imsets are useful because they can be used to **score** models consistently, and in particular can select the optimal **directed acyclic graph** model.

Outline

1. Imsets

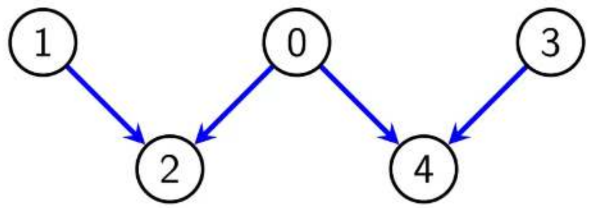
2. DAG Models

3. MAG Models

4. Model Search

DAG Models

Directed acyclic graphs (DAGs) can represent comparatively simple independence models.



We can use a **local Markov property** to completely define the model.

- pick a topological order;
- then each variable is conditionally independent of its predecessors in the ordering given its parents;

$$X_i \perp\!\!\!\perp X_{\mathrm{pre}(i)\setminus \mathrm{pa}(i)} \mid X_{\mathrm{pa}(i)}, \qquad \forall i \in V.$$

Imsets for DAG Models

Correspondingly, we can define the **standard imset** for a DAG $\mathcal{G}$ as:

$$u_{\mathcal{G}} := \sum_{i \in V} u_{\langle i, \mathrm{pre}(i) \mid \mathrm{pa}(i) \rangle}$$
$$= \delta_V - \delta_\emptyset + \sum_{i \in V} \left(\delta_{\mathrm{pa}(i)} - \delta_{\{i\} \cup \mathrm{pa}(i)} \right).$$

This has several nice properties:

- it is clearly a structural imset;
- P is Markov with respect to $\mathcal{G}$ if and only if $\langle \mathrm{I}_P, u_{\mathcal{G}} \rangle = 0$;

Imsets for DAG Models

Correspondingly, we can define the **standard imset** for a DAG $\mathcal{G}$ as:

$$\begin{aligned} u_{\mathcal{G}} &:= \sum_{i \in V} u_{\langle i, \mathrm{pre}(i) \mid \mathrm{pa}(i) \rangle} \\ &= \delta_V - \delta_\emptyset + \sum_{i \in V} \left(\delta_{\mathrm{pa}(i)} - \delta_{\{i\} \cup \mathrm{pa}(i)} \right). \end{aligned}$$

This has several nice properties:

- it is clearly a structural imset;
- P is Markov with respect to $\mathcal{G}$ if and only if $\langle \mathrm{I}_P, u_{\mathcal{G}} \rangle = 0$;
- $\mathcal{G}$ and $\mathcal{G}'$ are Markov equivalent if and only if $u_{\mathcal{G}} = u_{\mathcal{G}'}$;
- it is sparse (at most $2|V|$ terms).

Characteristic Imsets for DAG Models

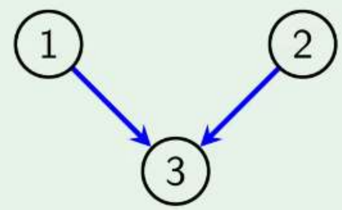
There is a bijective (Möbius) transformation we can make to obtain the **characteristic imset** (Studený et al., 2010) for a DAG:

$$c_{\mathcal{G}}(A) = 1 - \sum_{B \supseteq A} u_{\mathcal{G}}(B).$$

One can then show that

$$c_{\mathcal{G}}(A) = \begin{cases} 1 & \text{if } \exists v : \{v\} \subseteq A \subseteq \{v\} \cup \mathrm{pa}_{\mathcal{G}}(v) \\ 0 & \text{otherwise.} \end{cases}$$

Example. Consider the graph on the right. Then the non-zero sets are: $\emptyset$, $\{1\}$, $\{2\}$, $\{3\}$, $\{1,3\}$, $\{2,3\}$, $\{1,2,3\}$.

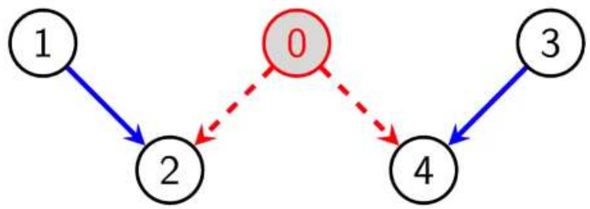


Outline

1. Imsets
2. DAG Models
3. MAG Models
4. Model Search

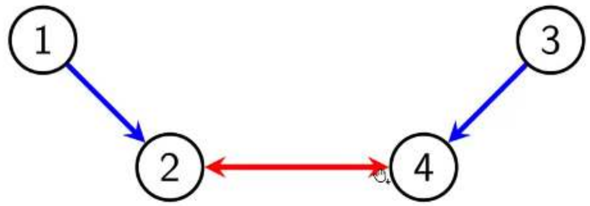
MAG Models

A (directed) **maximal ancestral graph** (MAG) model is just a collection of independences that can be represented by a DAG with hidden variables. (Richardson and Spirtes, 2002)



MAG Models

A (directed) **maximal ancestral graph** (MAG) model is just a collection of independences that can be represented by a DAG with hidden variables. (Richardson and Spirtes, 2002)



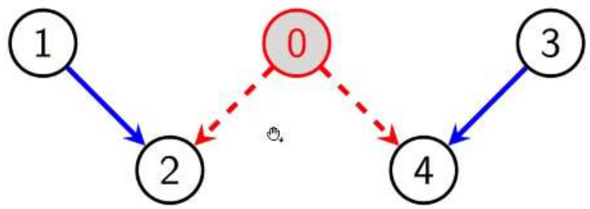
This MAG implies the independences

$$X_1 \perp\!\!\!\perp X_3, X_4 \qquad\qquad X_3 \perp\!\!\!\perp X_2 \mid X_1,$$

which cannot be faithfully represented by any DAG.

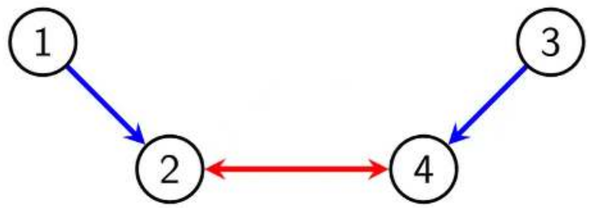
MAG Models

A (directed) **maximal ancestral graph** (MAG) model is just a collection of independences that can be represented by a DAG with hidden variables. (Richardson and Spirtes, 2002)



MAG Models

A (directed) **maximal ancestral graph** (MAG) model is just a collection of independences that can be represented by a DAG with hidden variables. (Richardson and Spirtes, 2002)



This MAG implies the independences

$$X_1 \perp\!\!\!\perp X_3, X_4 \qquad\qquad X_3 \perp\!\!\!\perp X_2 \mid X_1,$$

which cannot be faithfully represented by any DAG.

Markov Equivalence

In Hu and Evans (2020), we gave a criterion for two MAGs to be Markov equivalent based on collections of subsets.

Parametrizing Sets

The **parametrizing sets** for a MAG $\mathcal{G}$ are

$$\mathcal{S}(\mathcal{G}) = \{H \cup A : H \in \mathcal{H}(\mathcal{G}),\, A \subseteq \mathrm{tail}_{\mathcal{G}}(H)\},$$

where $\mathcal{H}(\mathcal{G})$ is the collection of **heads** in $\mathcal{G}$.

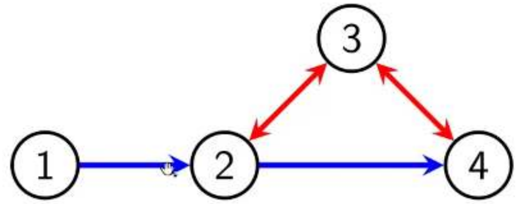
Given a vertex v in a head H, if we condition on $X_{H \setminus \{v\}}$, then the distribution cannot be m-separated from any $t \in \mathrm{tail}_{\mathcal{G}}(H)$.

As an analogy, for DAGs heads = vertices and tails = parent sets.

Parametrizing Sets Example

Consider this MAG, which implies

$$X_3 \perp\!\!\!\perp X_1 \quad \text{and} \quad X_4 \perp\!\!\!\perp X_1 \mid X_2 :$$

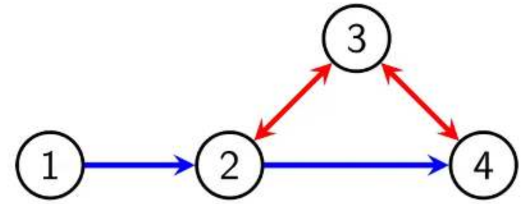


head	tail	parametrizing sets
$\{1\}$	$\emptyset$	
$\{2\}$	$\{1\}$	
$\{3\}$	$\emptyset$	
$\{2, 3\}$	$\{1\}$	
$\{4\}$	$\{2\}$	
$\{3, 4\}$	$\{1, 2\}$	

Parametrizing Sets Example

Consider this MAG, which implies

$$X_3 \perp\!\!\!\perp X_1 \quad \text{and} \quad X_4 \perp\!\!\!\perp X_1 \mid X_2 :$$



head	tail	parametrizing sets
$\{1\}$	$\emptyset$	$\{1\}$
$\{2\}$	$\{1\}$	$\{2\}, \{1,2\}$
$\{3\}$	$\emptyset$	$\{3\}$
$\{2,3\}$	$\{1\}$	$\{2,3\}, \{1,2,3\}$
$\{4\}$	$\{2\}$	$\{4\}, \{2,4\}$
$\{3,4\}$	$\{1,2\}$	$\{3,4\}, \{1,3,4\},$ $\{2,3,4\}, \{1,2,3,4\}$

Parametrizing set is missing only subsets $\{1,3\}$, $\{1,4\}$ and $\{1,2,4\}$.

Markov Equivalence Class and Characteristic Imsets

The parametrizing sets also give a representation of the Markov equivalence class of a MAG.

Theorem (Hu and Evans, 2020)

Two MAGs $\mathcal{G}$ and $\mathcal{G}'$ are Markov equivalent if and only if $\mathcal{S}(\mathcal{G}) = \mathcal{S}(\mathcal{G}')$.

Now note that the **characteristic imset** for a DAG takes the same form:

$$\begin{aligned} \mathcal{S}(\mathcal{G}) &= \{\{v\} \cup A : A \subseteq \mathrm{pa}_{\mathcal{G}}(v)\} \\ &= \{A : c_{\mathcal{G}}(A) = 1\}. \end{aligned}$$

So let's try using the parametrizing set to build the characteristic imset for a MAG!

Definition

Define the **characteristic imset** for a MAG $\mathcal{G}$ as

$$c_{\mathcal{G}}(A) = \begin{cases} 1 & \text{if } A \in \mathcal{S}(\mathcal{G}) \\ 0 & \text{otherwise.} \end{cases}$$

Retrofit for MAGs

Then define the **'standard' imset** as the inverse transformation of this.

$$u_{\mathcal{G}}(A) = \sum_{B \supseteq A} (-1)^{|B \setminus A|} (1 - c_{\mathcal{G}}(B)).$$

Proposition

Given a MAG $\mathcal{G}$, the 'standard' imset is the same as:

$$u_{\mathcal{G}} = \delta_V - \delta_{\emptyset} - \sum_{H \in \mathcal{H}(\mathcal{G})} \sum_{W \subseteq H} (-1)^{|H \setminus W|} \delta_{W \cup T},$$

where $T = \mathrm{tail}_{\mathcal{G}}(H)$.

Note that this is consistent with the definition for DAGs.

Defining the Model

We used ILPs to check which 'standard' imsets define the model.

There are three cases, based on whether the 'standard' imset $u_{\mathcal{G}}$:

(i) does define the **same model** as $\mathcal{G}$;

(ii) defines a model with a (strict) **subset** of the independence restrictions of $\mathcal{G}$;

(iii) is **not structural** (so does not define any model).

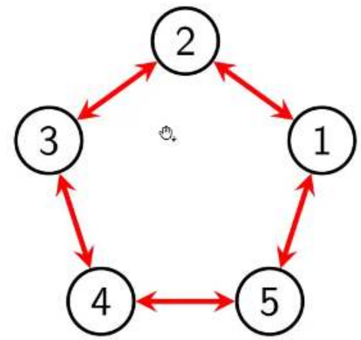
For small graphs, we find that they *usually* fall into category (i).

For all MAGs with 5 or 6 nodes, and 7 nodes and ≤ 13 or ≥ 18 edges:

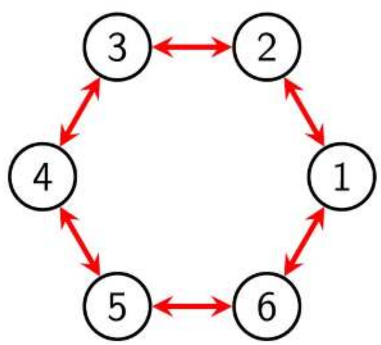
n	equiv. classes	(i)	(ii)	(iii)
5	285	284	1	0
6	13,303	13,248	54	1
7^*	1,161,461	1,146,501	14,562	8

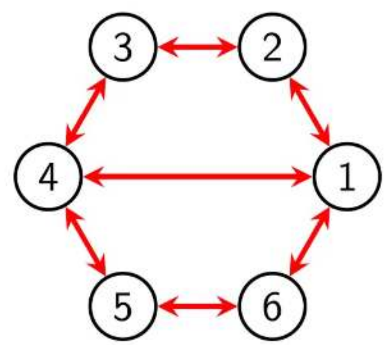
Defining the Model

For $n = 5$ 'standard' imsets **all** define the model, **except** for the bidirected 5-cycle.



The bidirected 6-cycle is not even **structural**.





'Simple' MAGs

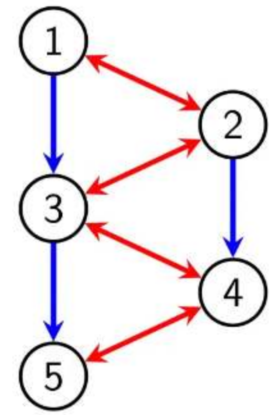
Definition

We say that a MAG is **simple** if its maximal head size is at most two.

$\|V\|$	equiv. classes	simple MAGs	DAGs
5	285	205	119
6	13,303	6,278	2,025
7*	1,161,461	331,310	57,661

*having at most 13 or at least 18 edges.

Example.



'Simple' MAGs

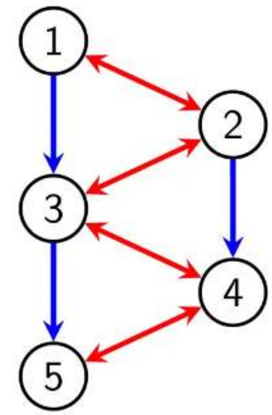
Definition

We say that a MAG is **simple** if its maximal head size is at most two.

$\lvert V \rvert$	equiv. classes	simple MAGs	DAGs
5	285	205	119
6	13,303	6,278	2,025
7*	1,161,461	331,310	57,661

*having at most 13 or at least 18 edges.

Example.



Proposition

For every simple MAG $\mathcal{G}$, the standard imset **does** define the model implied by the graph. In addition, it contains at most $2(|V| + |E|)$ terms.

Model Scoring

Usual consistent score for model scoring is the BIC. This requires us to find the maximum likelihood for each model we score.

We have a proposal for a scoring models (Andrews, 2022):

$$h(\mathcal{G}) := 2n\langle \mathrm{I}_P, u_\mathcal{G} \rangle - k \log n,$$

where n is the number of samples, k is the number of parameters, and I_P is the **interaction information operator** (see appendix).

If $u_\mathcal{G}$ defines the model, we have

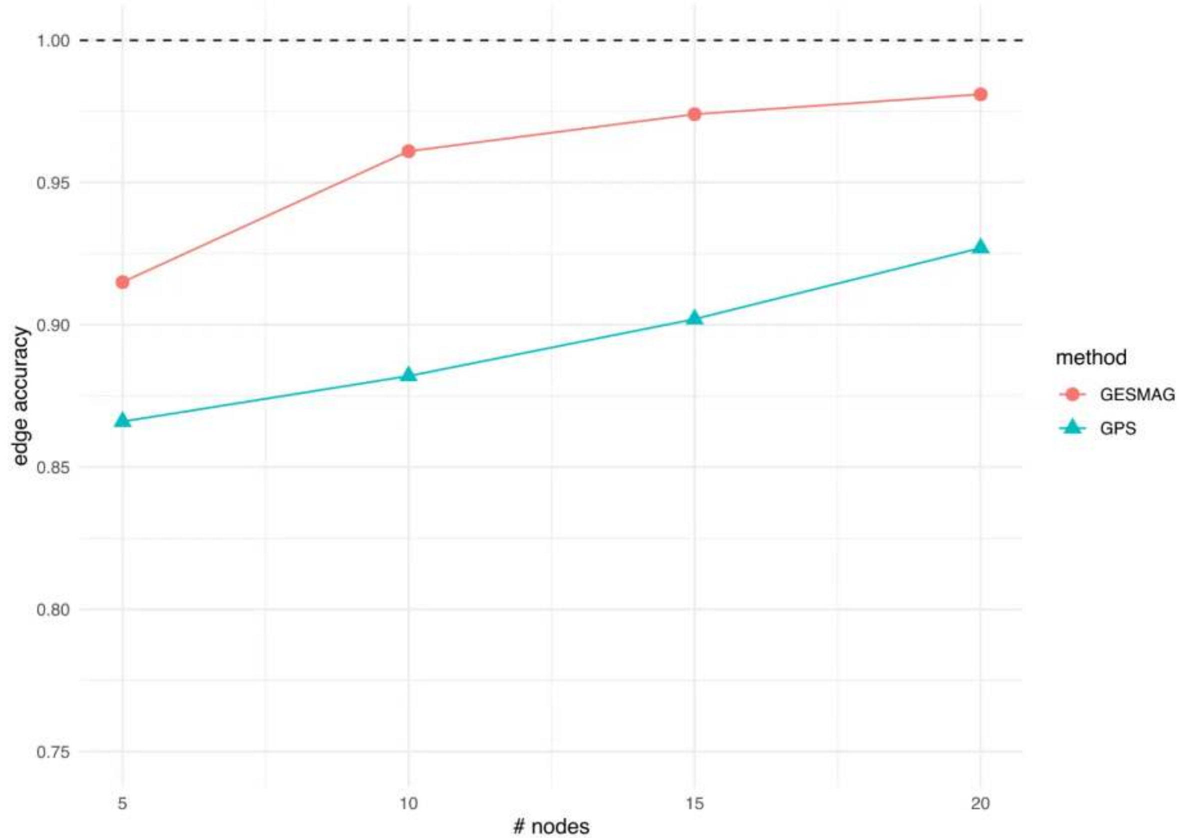
$$n\langle \mathrm{I}_P, u_\mathcal{G} \rangle \approx \ell_\mathcal{G}(P; \boldsymbol{X}_V),$$

so our score approximates the BIC.

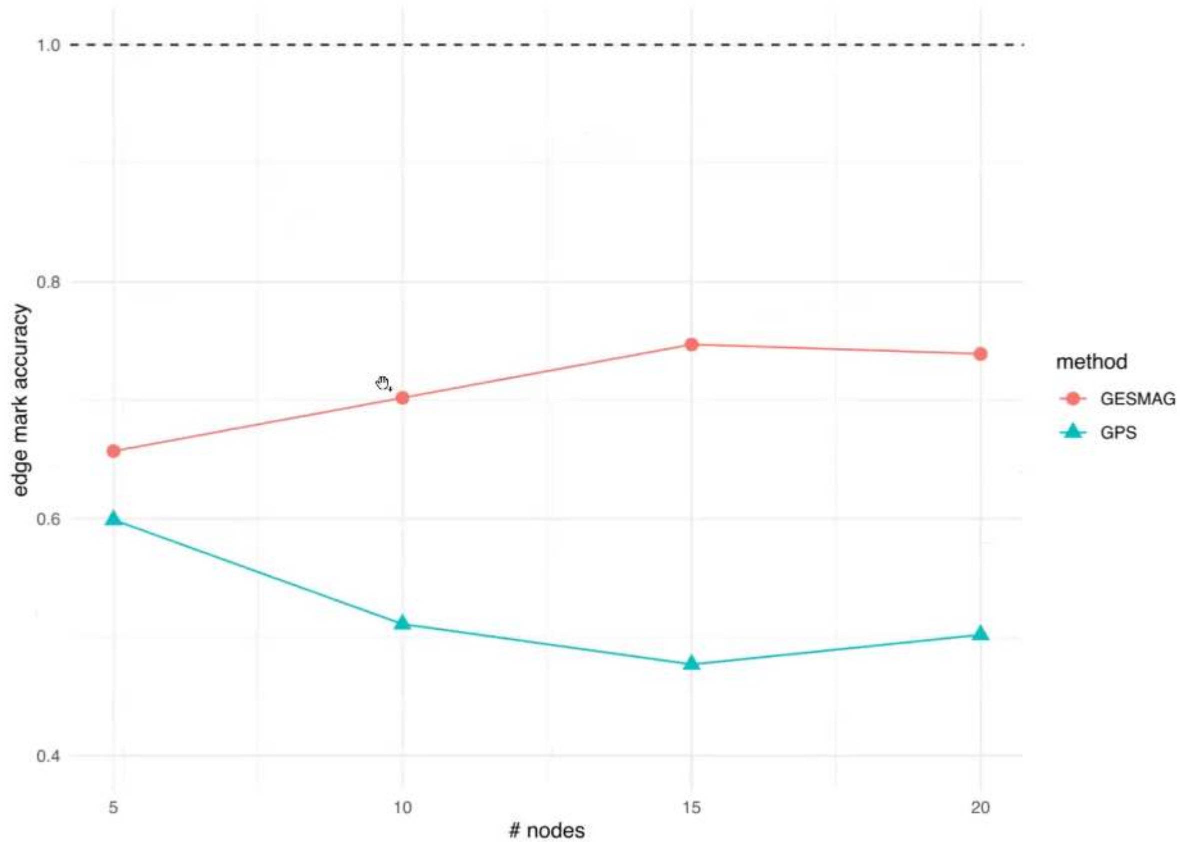
Hence the score is consistent over this set of MAGs
(i.e. the highest score is given asymptotically to the true model).

Since $u_\mathcal{G}$ ought to define the model, we restrict to simple MAGs.

Results (Skeleton accuracy)



Results (Edge mark accuracy)



Summary

- Imsets can be used to define arbitrary conditional independence models;
- they have particularly nice properties when applied to DAGs.
- *Some* of those properties are replicated in MAGs, but (unfortunately) not all of them!
- Problem is that (for some graphs) it is not possible to describe the model without using conditional independences that repeat sets.

  Describing these graphs is an **open problem**, and seems to be combinatorially difficult.

Summary

- Imsets can be used to define arbitrary conditional independence models;
- they have particularly nice properties when applied to DAGs.
- *Some* of those properties are replicated in MAGs, but (unfortunately) not all of them!
- Problem is that (for some graphs) it is not possible to describe the model without using conditional independences that repeat sets.

  Describing these graphs is an **open problem**, and seems to be combinatorially difficult.
- Imsets that **do** represent the models can be used to give a new consistent score, which is easier to compute than the BIC.
- We have also developed a greedy algorithm for quickly learning simple MAGs.

References I

Andrews—*Inducing Sets: A New Perspective for Ancestral Graph Markov Models*, PhD Thesis, U. Pitt, 2022.

Andrews et al.—The m-connecting imset and factorization for ADMG models, *arXiv:2207.08963*, 2022.

Bouckaert et al.—Efficient algorithms for conditional independence inference, *JMLR*, 2010.

Claassen and Bucur—Greedy equivalence search in the presence of latent confounders, *UAI*, 2022.

Hemmecke et al.—Characteristic imsets for learning Bayesian network structure. *IJAR*, 2012.

Hu and Evans— Faster algorithms for Markov equivalence, *UAI-20*, 2020.

Hu and Evans—Towards standard imsets for maximal ancestral graphs. *arXiv:2208.10436*, 2022.

Thank you!