Expressive Power of Graphical Models

Michael Gutmann

Probabilistic Modelling and Reasoning (INFR11134)
School of Informatics, University of Edinburgh

Spring Semester 2019

Recap

- Need for efficient representation of probabilistic models
  - Restrict the number of directly interacting variables by making independence assumptions
  - Restrict the form of interaction by making parametric family assumptions.
- Directed and undirected graphs to represent independencies (I-maps)
- Equivalences between independencies (Markov properties) and factorisation
- Rules for reading independencies from the graph that hold for all distributions that factorise over the graph.

Program

1. Minimal independency maps
2. (Lossy) conversion between directed and undirected I-maps

Program

1. Minimal independency maps
   - Definition of minimal I-maps, the goal of a perfect maps
   - Construction of undirected minimal I-maps and their uniqueness
   - Construction of directed minimal I-maps and their non-uniqueness
   - Equivalence of I-maps (I-equivalence)

2. (Lossy) conversion between directed and undirected I-maps

I-maps

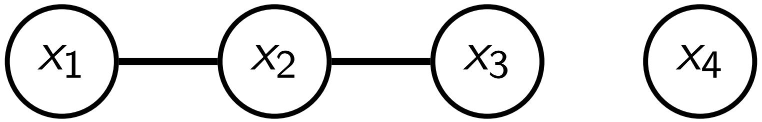
- A graph is an independency map (I-map) for a set of independencies $\mathcal{I}$ if the independencies asserted by the graph are part of $\mathcal{I}$.
- Criterion for an I-map is that the independency assertions are true.
- Is not concerned with the number of independency assertions.
- Drawback of I-maps: they may not be very useful because they may"miss" many independencies of $\mathcal{I}$.
- Full graph does not make any assertions. Empty set is trivially a subset of $\mathcal{I}$, so that the full graph is trivially an I-map.

Minimal I-maps

- Minimal I-map : graph such that if you remove an edge (more independencies), the graph is not an I-map any more.
- Intuitively, the point of minimal I-maps is to "sparsify" I-maps so that they become more useful (note: while sparser, the independence assertions must still be correct for a graph to be an I-map)
- Generally, we want the graph to represent as many true independencies as possible: graph is sparser, and thus more informative, easier to understand, and facilitates learning and inference.
- If the graph represents all independencies in $\mathcal{I}$, the graph is said to be a perfect map (P-map).
  (May be hard to find and will not always exist!)

Example

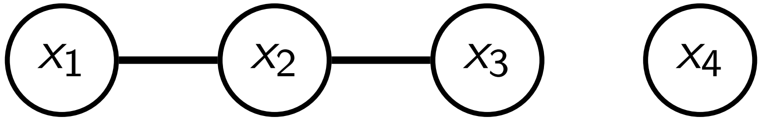
- Let $p(x_1, x_2, x_3, x_4) \propto \phi_1(x_1, x_2)\phi_2(x_2, x_3)\phi_3(x_4)$
- Denote the set of independencies that hold for p by $\mathcal{I}(p)$
- Minimal I-map for $\mathcal{I}(p)$:



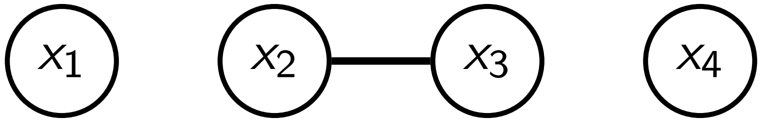
- I-map because p factorises over the graph and hence all independencies asserted by the graph must hold for p.
- Minimal I-map because removing an edge results in a graph that makes wrong independency assertions.

Example

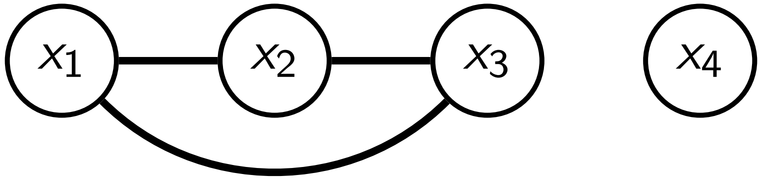
- Let $p(x_1, x_2, x_3, x_4) \propto \phi_1(x_1, x_2)\phi_2(x_2, x_3)\phi_3(x_4)$
- Minimal I-map for $\mathcal{I}(p)$:



- Not an I-map for $\mathcal{I}(p)$ (wrongly claims $x_1 \perp\!\!\!\perp \{x_2, x_3\}$):

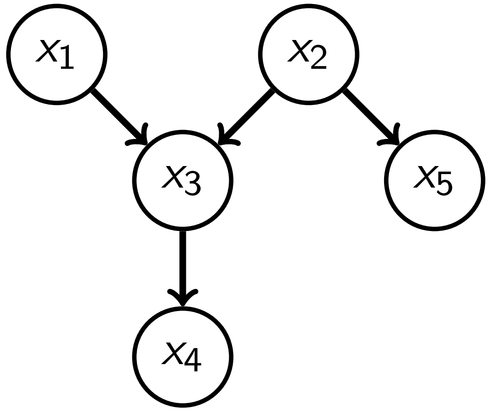


- (Non-minimal) I-map for $\mathcal{I}(p)$ ($x_1 - x_3$ edge could be removed):



Example

- Let $p(x_1, x_2, x_3, x_4, x_5) = p(x_1)p(x_2)p(x_3|x_1, x_2)p(x_4|x_3)p(x_5|x_2)$
- Denote the set of independencies that hold for p by $\mathcal{I}(p)$
- Minimal I-map for $\mathcal{I}(p)$:

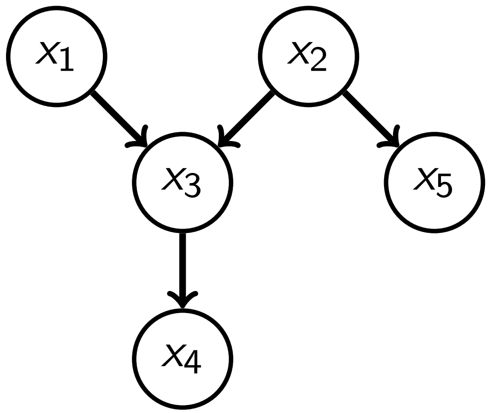


- I-map because p factorises over the graph and hence all independencies asserted by the graph must hold for p.
- Minimal I-map because removing an edge results in a graph that makes wrong independency assertions.

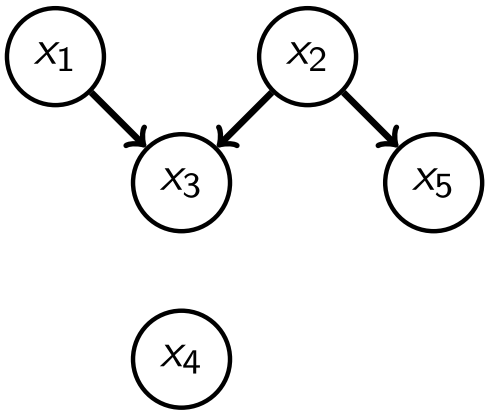
Example

Let $p(x_1, x_2, x_3, x_4, x_5) = p(x_1)p(x_2)p(x_3|x_1, x_2)p(x_4|x_3)p(x_5|x_2)$

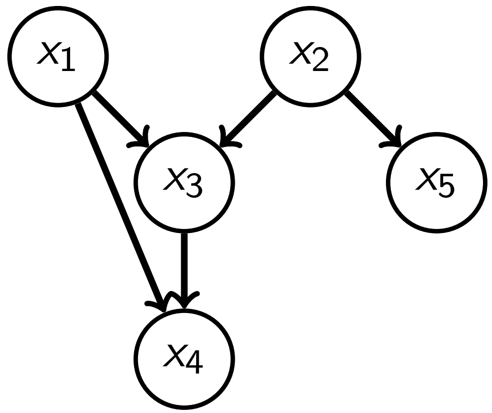
Minimal I-map for $\mathcal{I}(p)$:



Not an I-map for $\mathcal{I}(p)$:
(wrongly claims $x_4 \perp\!\!\!\perp x_3$)



(Non-minimal) I-map for $\mathcal{I}(p)$:
($x_1 \to x_4$ could be removed)



Constructing undirected minimal I-maps

Given random variables $\mathbf{x} = (x_1, \ldots, x_d)$ with positive distribution $p > 0$

- Approaches based on pairwise and local Markov property
- Both yield same (unique) graph.
- For local Markov property approach: For each node:
  1. determine its Markov blanket $\mathrm{MB}(x_i)$:
     minimal set of nodes U such that

     $$x_i \perp\!\!\!\perp \{\text{all variables} \setminus (x_i \cup U)\} \mid U$$

     with respect to p.
  2. we know that x_i and $\mathrm{MB}(x_i)$ must be neighbours in the graph:
     Connect x_i to all nodes in $\mathrm{MB}(x_i)$
- We need $p > 0$ because otherwise local independencies may not imply global ones (see slides on undirected graphical models).

Constructing directed minimal I-maps

Given a distribution p.

- We can use the ordered Markov property to derive a directed graph that is a minimal I-map for $\mathcal{I}(p)$.

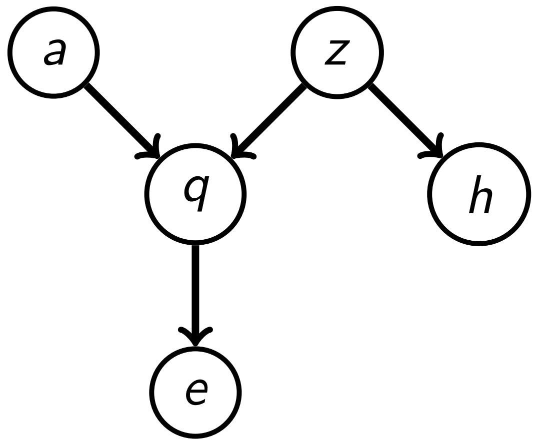
$$x_i \perp\!\!\!\perp \{\mathrm{pre}_i \setminus \mathrm{pa}_i\} \mid \mathrm{pa}_i$$

- Procedure is exactly the same as the one used to simplify the factorisation obtained by the chain rule:
  1. Assume an ordering of the variables. Denote the ordered random variables by $x_1, \ldots, x_d$.
  2. For each i, find a minimal subset of variables $\pi_i \subseteq \mathrm{pre}_i$ such that
  $$x_i \perp\!\!\!\perp \{\mathrm{pre}_i \setminus \pi_i\} \mid \pi_i$$
  holds in $\mathcal{I}(p)$.
  3. Construct a graph with parents $\mathrm{pa}_i = \pi_i$.

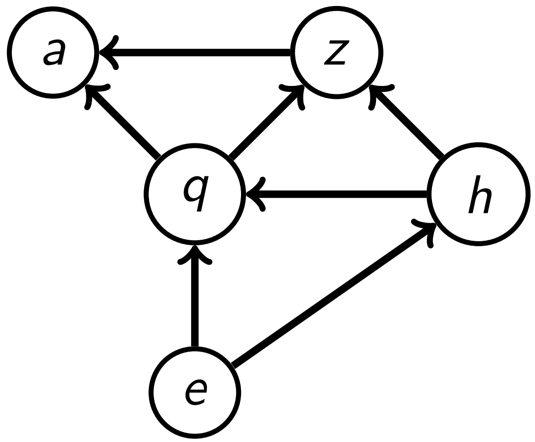
Directed minimal I-maps are not unique

Consider $p(a, z, q, e, h) = p(a)p(z)p(q|a, z)p(e|q)p(h|z)$

For ordering (a, z, q, e, h)



For ordering (e, h, q, z, a)



- Directed (minimal) I-maps are not unique
- Different directed (minimal) I-maps for the same p may not make the same independence assertions.
- Minimal I-maps of $\mathcal{I}(p)$ may not represent all independencies that hold for p, but generally only a subset of them.

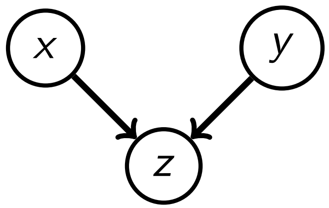
I-equivalence for directed graphs

- How do we determine whether two directed graphs make the same independence assertions (that they are "I-equivalent")?
- From d-separation: what matters is
  - which node is connected to which irrespective of direction (skeleton)
  - the set of collider (head-to-head) connections

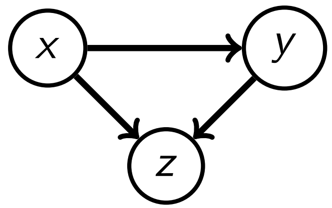
Connection	$p(x, y)$	$p(x, y \mid z)$
$x \rightarrow z \rightarrow y$	$x \not\perp\!\!\!\perp y$	$x \perp\!\!\!\perp y \mid z$
$x \leftarrow z \rightarrow y$	$x \not\perp\!\!\!\perp y$	$x \perp\!\!\!\perp y \mid z$
$x \rightarrow z \leftarrow y$	$x \perp\!\!\!\perp y$	$x \not\perp\!\!\!\perp y \mid z$

I-equivalence for directed graphs

- The situation $x \perp\!\!\!\perp y$ and $x \not\perp\!\!\!\perp y \mid z$ can only happen if we have colliders without "covering edge" $x \rightarrow y$ or $x \leftarrow y$
- Colliders without covering edge are called "immoralities"
- Theorem: For two directed graphs G_1 and G_2:
  $G1$ and G_2 are I-equivalent $\Longleftrightarrow$ G_1 and G_2 have the same skeleton and the same set of immoralities.



$x \perp\!\!\!\perp y$ and $x \not\perp\!\!\!\perp y \mid z$
Collider w/o covering edge

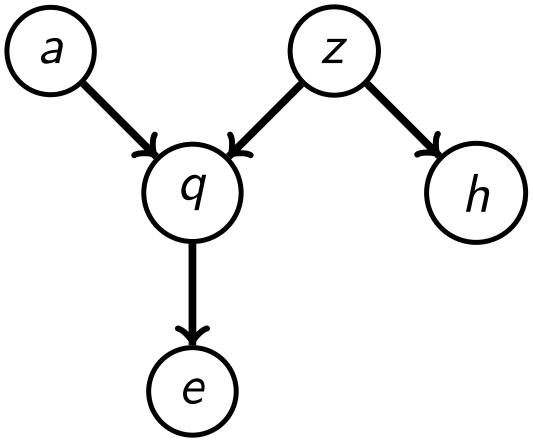


$x \not\perp\!\!\!\perp y$ and $x \not\perp\!\!\!\perp y \mid z$
Collider with covering edge

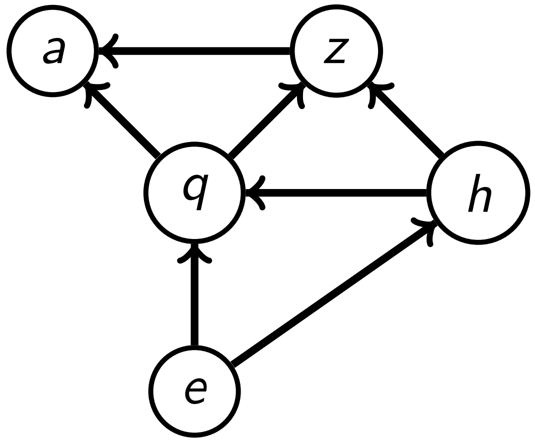
Example

Not I-equivalent because of skeleton mismatch:

G_1:



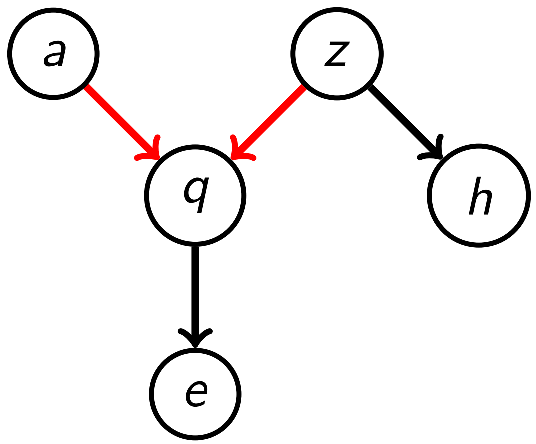
G_2:



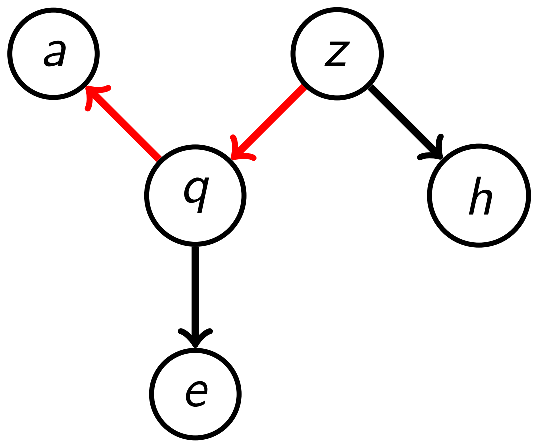
Example

Not I-equivalent because of immoralities mismatch:

G_1:



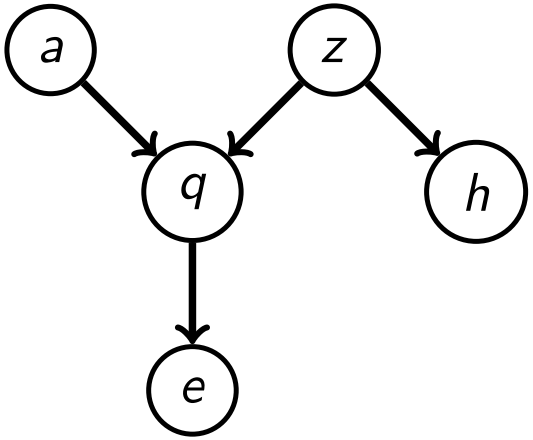
G_2:



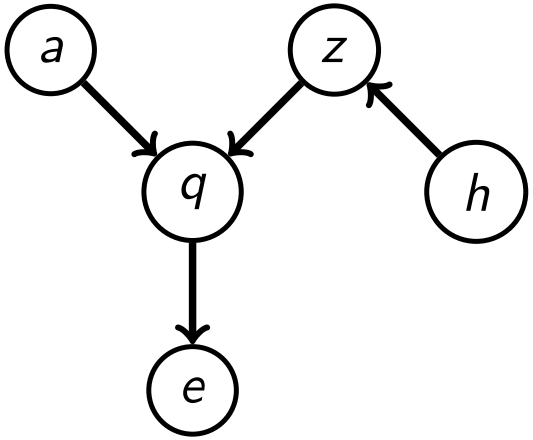
Example

I-equivalent (same skeleton, same immoralities):

G_1:



G_2:



I-equivalence for undirected graphs?

- For undirected graphs, minimal I-map is unique.
- Different graphs make different independence assertions.
- Equivalence question does not come up.

Program

1. Minimal independency maps

2. (Lossy) conversion between directed and undirected I-maps
   - Moralisation for directed $\rightarrow$ undirected I-map
   - Example of non-existence of undirected perfect map
   - Triangulation for undirected $\rightarrow$ directed I-map
   - Example of non-existence of directed perfect map
   - Strengths and weaknesses of directed and undirected graphical models

Directed to undirected graphical model

Goal: undirected minimal I-Map. Assume directed I-map G given

- Probabilistic models factorises according to G as

$$p(x_1, \ldots, x_d) = \prod_{i=1}^{d} p(x_i | \mathrm{pa}_i)$$

- Write each $p(x_i|\mathrm{pa}_i)$ as factor $\phi_i(x_i, \mathrm{pa}_i)$:

$$p(x_1, \ldots, x_d) = \prod_{i=1}^{d} \phi_i(x_i, \mathrm{pa}_i)$$

  Gibbs distribution with normalisation constant equal to one
- Graph operation: Form cliques for (x_i, pa_i)

Directed to undirected graphical model

Goal: undirected minimal I-Map. Assume directed I-map G given

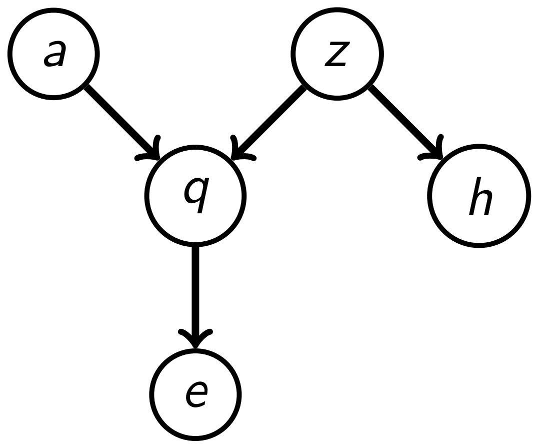
$$p(x_1, \ldots, x_d) = \prod_{i=1}^{d} p(x_i|\mathrm{pa}_i) = \prod_{i=1}^{d} \phi_i(x_i, \mathrm{pa}_i)$$

- Graph operation: Form cliques for (x_i, pa_i)
- Remove arrows, and add edges between *all* parents of x_i.
- Conversion from directed to undirected graphical model is called "moralisation". Obtained undirected graph is the "moral graph" of G.
- Process above is equivalent to using the directed graph to determine the Markov blanket for each x_i.

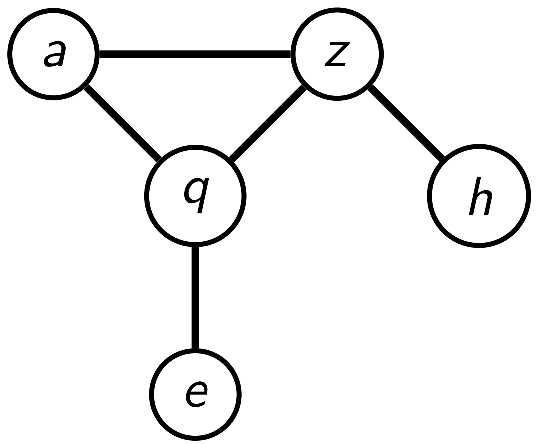
Example

Goal: Undirected minimal I-map for
$p(a, z, q, e, h) = p(a)p(z)p(q|a, z)p(e|q)p(h|z)$

Given: directed I-map



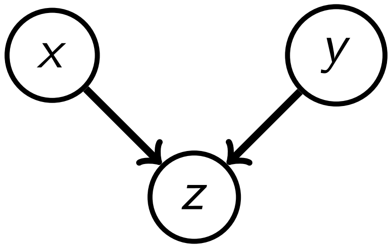
Moral graph:



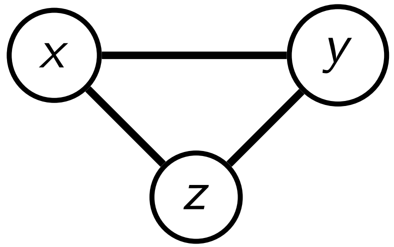
Note: In the undirected I-map, we do not have $a \perp\!\!\!\perp z$. We lost that information.
Minimal I-maps of $\mathcal{I}(p)$ may not represent all independencies that hold for p, but generally only a subset of them.

Simpler example

Goal: Undirected minimal I-map for $p(x, y, z) = p(x)p(y)p(z|x, y)$



Given: directed I-map



Only possible undirected I-map is full graph

There is no undirected I-map representing $\mathcal{I} = \{x \perp\!\!\!\perp y,\; x \not\perp\!\!\!\perp y \mid z\}$

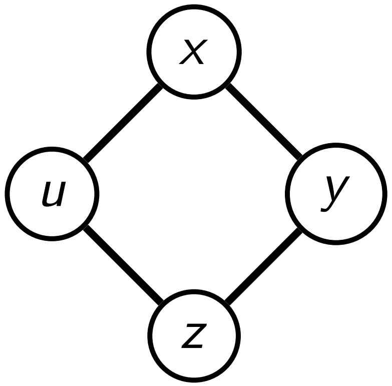
Undirected to directed graphical model

Goal: directed minimal I-Map. Assume undirected I-map H given

- We can use the approach based on the local Markov property
- Read required independencies from the undirected graph
- Typically results in directed graphs that are larger than the undirected graph
- Directed graph will not have any immoralities
  (for proof, see e.g. theorem 4.10 in Koller and Friedman's book, not examinable)
- Results in chordal/triangulated graphs (longest loop without shortcuts is a triangle).

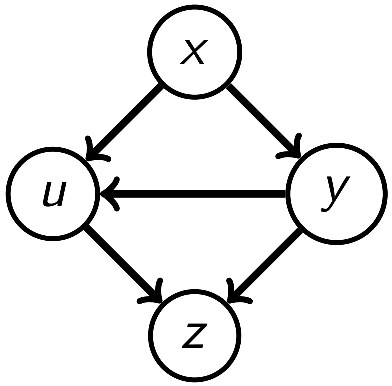
Example

Goal: Directed minimal I-map for
$p(x, y, z, u) \propto \phi_1(x, y)\phi_2(y, z)\phi_3(z, u)\phi_4(u, x)$



Given: undirected I-map

$$x \perp\!\!\!\perp z \mid u, y$$
$$u \perp\!\!\!\perp y \mid x, z$$



Directed minimal I-map
(with ordering: x, y, u, z)

$$x \perp\!\!\!\perp z \mid u, y$$
$$u \not\!\perp\!\!\!\perp y \mid x, z$$

We lost information with the conversion. There is no directed I-map representing $\mathcal{I} = \{x \perp\!\!\!\perp z \mid u, y,\ u \perp\!\!\!\perp y \mid x, z\}$

Strengths and weaknesses

- Both directed and undirected graphical models have strengths and weaknesses
- Some independencies are more easily represented with directed graphs, others with undirected graphs.
- Undirected graphs are suitable when interactions are symmetrical and when there is no natural ordering of the variables, but they cannot represent “explaining away” scenario (colliders).
- Directed graphs are suitable when we have an idea of the data generating process (e.g. what is “causing” what, ancestral sampling), but they may force directionality where there is none, yielding unintuitive graphs (see triangulation).
- It is possible to combine individual strengths with mixed/partially directed graphs.

Program recap

1. Minimal independency maps
   - Definition of minimal I-maps, the goal of a perfect maps
   - Construction of undirected minimal I-maps and their uniqueness
   - Construction of directed minimal I-maps and their non-uniqueness
   - Equivalence of I-maps (I-equivalence)
2. (Lossy) conversion between directed and undirected I-maps
   - Moralisation for directed $\to$ undirected I-map
   - Example of non-existence of undirected perfect map
   - Triangulation for undirected $\to$ directed I-map
   - Example of non-existence of directed perfect map
   - Strengths and weaknesses of directed and undirected graphical models