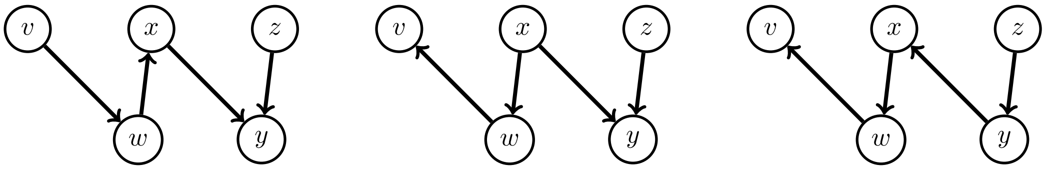
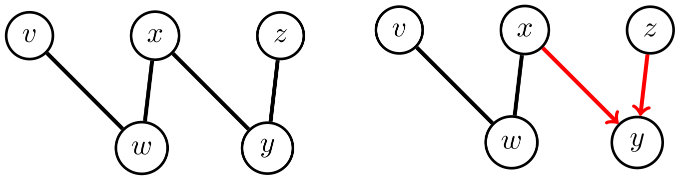
Exercise 1. *I-maps*

(a) *Which of three graphs represent the same set of independencies? Explain.*



Graph 1 *Graph 2* *Graph 3*

Solution. To check whether the graphs are I-equivalent, we have to check the skeletons and the immoralities. All have the same skeleton, but graph 1 and graph 2 also have the same immorality. The answer is thus: graph 1 and 2 encode the same independencies.



skeleton immorality

(b) *For $p(a, z, q, e, h) = p(a)p(z)p(q|a, z)p(e|q)p(h|z)$, determine a minimal I-map for the orderings*

- (a, z, q, e, h)
- (a, z, h, q, e)
- (e, h, q, z, a)

Are the I-maps I-equivalent?

Solution. To find a minimal I-map, we can use the same procedure we 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 $\pi_i \subseteq \text{pre}_i$ such that

$$x_i \perp\!\!\!\perp \text{pre}_i \setminus \pi_i \mid \pi_i$$

holds for p.

3. Construct a graph with parents $\text{pa}_i = \pi_i$.

Checking whether the independencies hold can be difficult. But here, we are given a factorisation for p from which we can construct the directed graph in Figure 1. We thus can use graphical methods to check whether an independency holds (e.g. by d-separation).

Note: If the graph does not indicate that a certain independency holds, we had to generally check, however, whether it indeed does not hold for a specific distribution. If we don't, we won't obtain a minimal I-map but just an I-map. This is because the graph may not be a perfect map, and p may have independencies that are not encoded in the graph. Here, this won't be needed because we left the conditionals of p unspecified. In other words, this means that we here construct minimal I-maps for the independencies that hold for all $p(a, z, q, e, h)$ that factorise as $p(a, z, q, e, h) = p(a)p(z)p(q|a, z)p(e|q)p(h|z)$.

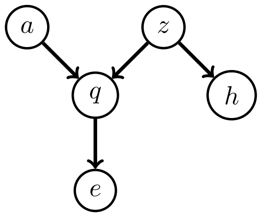


Figure 1: Exercise 1, question (b): I-map for the orderings (a, z, h, q, e) and (a, z, q, e, h).

The graph in Figure 1 is exactly the I-map for the ordering (a, z, q, e, h).

The ordering (a, z, h, q, e) gives rise to the same graph.

For the ordering (e, h, q, z, a), we build a graph where e is the root. From Figure 1, we see that $h \perp\!\!\!\perp e$ does not hold. We thus set e as parent of h, see first graph in Figure 2. Then:

- We consider q: $\text{pre}_q = \{e, h\}$. There is no subset π_q of pre_q on which we could condition to make q independent of $\text{pre}_q \setminus \pi_q$, so that we set the parents of q in the graph to $\text{pa}_q = \{e, h\}$. (Second graph in Figure 2.)
- We consider z: $\text{pre}_z = \{e, h, q\}$. From the graph in Figure 1, we see that for $\pi_z = \{q, h\}$ we have $z \perp\!\!\!\perp \text{pre}_z \setminus \pi_z | \pi_z$. Note that $\pi_z = \{q\}$ does not work because $z \perp\!\!\!\perp e, h|q$ does not hold. We thus set $\text{pa}_z = \{q, h\}$. (Third graph in Figure 2.)
- We consider a: $\text{pre}_a = \{e, h, q, z\}$. This is the last node in the ordering. To find the minimal set π_a for which $a \perp\!\!\!\perp \text{pre}_a \setminus \pi_a | \pi_a$, we can determine its Markov blanket $\text{MB}(a)$. The Markov blanket is the set of parents (none), children (q), and co-parents of a (z) in Figure 1, so that $\text{MB}(a) = \{q, z\}$. We thus set $\text{pa}_a = \{q, z\}$.(Fourth graph in Figure 2.)

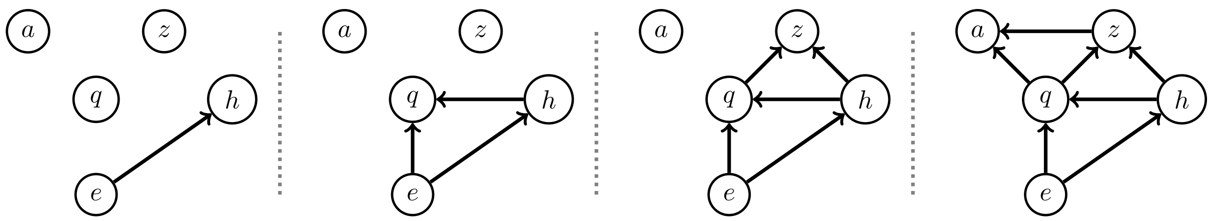


Figure 2: Exercise 1, Question (b):Construction of a minimal directed I-map for the ordering (e, h, q, z, a).

We thus see that the orderings (a, z, q, e, h) and (a, z, h, q, e) give I-equivalent minimal I-maps (Figure 1) while (e, h, q, z, a) yields a denser graph (Figure 2) that is not I-equivalent. While a minimal I-map it does e.g. not show that $a \perp\!\!\!\perp z$.

(c) *For the collection of random variables (a, z, h, q, e) you are given the following Markov blankets for each variable:*

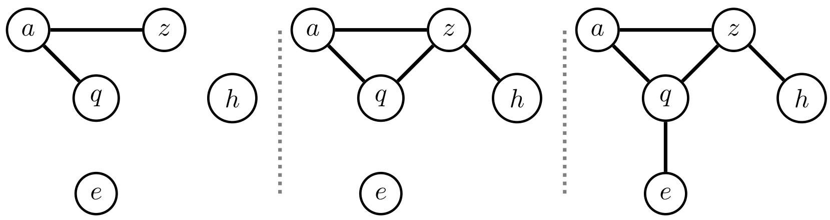
- *MB(a) = $\{q,z\}$*

- *MB(z)* = $\{a,q,h\}$
- *MB(h)* = $\{z\}$
- *MB(q)* = $\{a,z,e\}$
- *MB(e)* = $\{q\}$

(i) Draw the undirected minimal I-map.

(ii) Indicate a Gibbs distribution that satisfies the independence relations specified by the Markov blankets.

Solution. Connecting each variable to all variables in its Markov blanket yields the desired undirected minimal I-map (see lecture slides). Note that the Markov blankets are not mutually disjoint.



After MB(a) After MB(z) After MB(q)

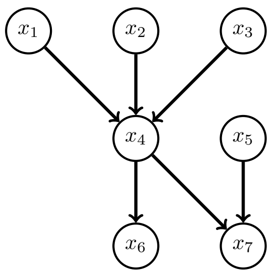
For positive distributions, the set of distributions that satisfy the local Markov property relative to a graph (as given by the Markov blankets) is the same as the set of Gibbs distributions that factorise according to the graph. Given the I-map, we can now easily find the Gibbs distribution

$$p(a, z, h, q, e) = \phi_1(a, z, q)\phi_2(q, e)\phi_3(z, h)$$

Note that we used the maximal clique (a, z, q).

Exercise 2. *Conversion between graphs*

(a) For distributions that factorises over the graph below, find the minimal undirected I-map.

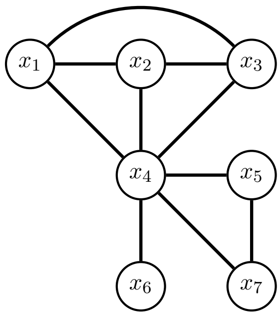


Solution. To derive an undirected minimal I-map from a directed one, we have to construct the moralised graph where the "unmarried" parents are connected by a covering edge. This is because each conditional $p(x_i|\mathrm{pa}_i)$ corresponds to a factor $\phi_i(x_i, \mathrm{pa}_i)$ and we need to connect all variables that are arguments of the same factor with edges.

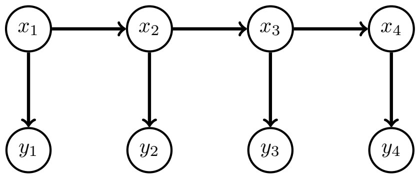
Statistically, the reason for marrying the parents is as follows: An independency $x \perp\!\!\!\perp y|\{\text{child, other nodes}\}$ does not hold in the directed graph in case of collider connections but would hold in the undirected graph if we didn't marry the parents. Hence links between the parents must be added.

It is important to add edges between *all* parents of a node. Here, $p(x_4|x_1, x_2, x_3)$ corresponds to a factor $\phi(x_4, x_1, x_2, x_3)$ so that all four variables need to be connected. Just adding edges $x_1 - x_2$ and $x_2 - x_3$ is not enough.

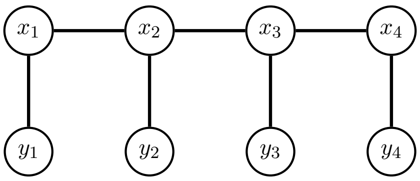
The moral graph, which is the requested minimal undirected I-map, is shown below.



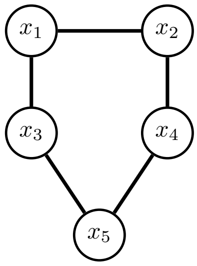
(b) *The graph below is a directed minimal I-map for the hidden Markov model. Find the corresponding undirected minimal I-map.*



Solution. The graph does not contain any head-head connections. The undirected minimal I-map is thus obtained by removing all arrows from the graph.



(c) *For the undirected I-map below, what is a corresponding directed minimal I-map?*



Solution. We use the ordering x_1, x_2, x_3, x_4, x_5 and follow the general procedure to construct the directed minimal I-map while reading the independencies from the undirected graph:

- x_2 is not independent from x_1 so that we set $\text{pa}_2 = \{x_1\}$. See first graph in Figure 3.
- Since x_3 is connected to both x_1 and x_2, we generally don't have $x_3 \perp\!\!\!\perp x_2, x_1$. We cannot make x_3 independent from x_2 by conditioning on x_1 because there are two paths from x_3 to x_2 and x_1 only blocks the upper one. Moreover, x_1 is a neighbour of x_3 so that conditioning on x_2 does make them independent. Hence we must set $\text{pa}_3 = \{x_1, x_2\}$. See second graph in Figure 3.
- For x_4, we see from the undirected graph, that $x_4 \perp\!\!\!\perp x_1 \mid x_3, x_2$. The graph further shows that removing either x_3 or x_2 from the conditioning set is not possible and conditioning on x_1 won't make x_4 independent from x_2 or x_3. We thus have $\text{pa}_4 = \{x_2, x_3\}$. See fourth graph in Figure 3.
- The same reasoning shows that $\text{pa}_5 = \{x_3, x_4\}$. See last graph in Figure 3.

This results in the triangulated directed graph in Figure 3 on the right.

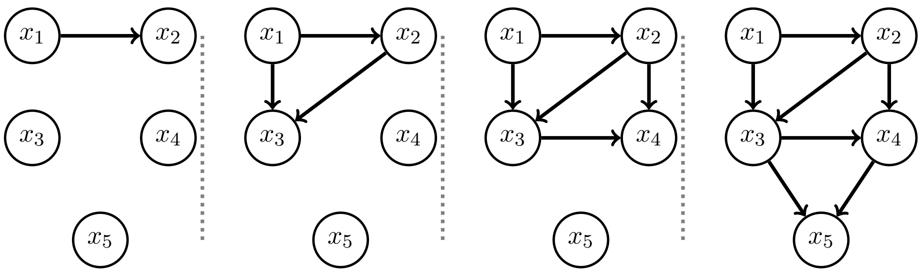


Figure 3: . Answer to Exercise 2, Question (c).

To see why triangulation is necessary consider the case where we didn't have the edge between x_2 and x_3 as in Figure 4. The directed graph would then imply that $x_3 \perp\!\!\!\perp x_2 \mid x_1$ (check!). But this independency assertion does not hold in the undirected graph so that the graph in Figure 4 is not an I-map.

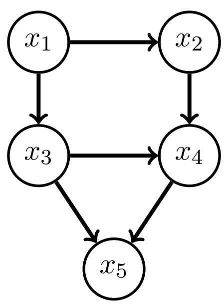
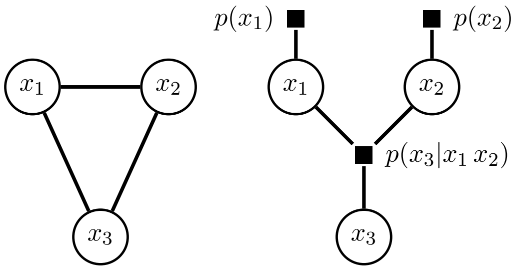


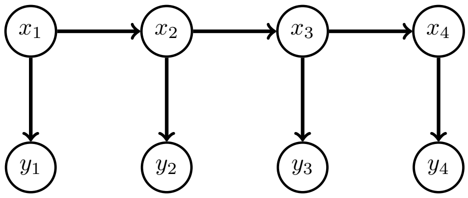
Figure 4: Not a directed I-map for the undirected graphical model defined by the graph in Question(c) of Exercise 2.

(d) Draw an undirected graph and an undirected factor graph for $p(x_1, x_2, x_3) = p(x_1)p(x_2)p(x_3|x_1, x_2)$

Solution.



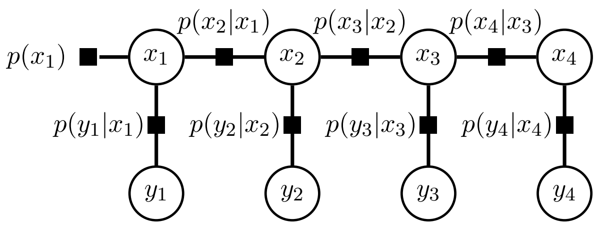
(e) Draw an undirected factor graph for the directed graphical model defined by the graph below.



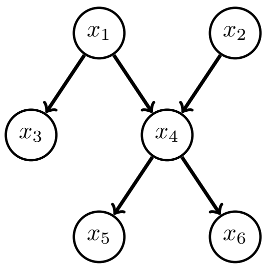
Solution. The graph specifies probabilistic models that factorise as

$$p(x_1, \ldots, x_4, y_1, \ldots, y_4) = p(x_1)p(y_1|x_1) \prod_{i=2}^{4} p(y_i|x_i)p(x_i|x_{i-1})$$

It is the graph for a Hidden Markov model. The corresponding factor graph is shown below.



(f) Draw the moralised graph and an undirected factor graph for directed graphical models defined by the graph below (this kind of graph is called a polytree: there are no loops but a node may have more than one parent).

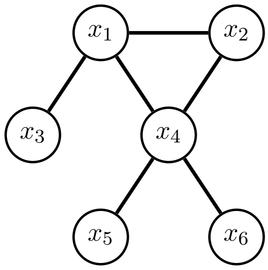


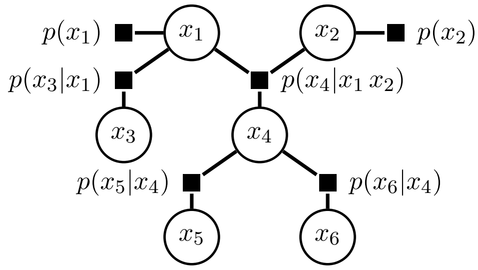
Solution. The moral graph is obtained by connecting the parents of the collider node x_4. See the graph on the left in the figure below.

For the factor graph, we note that the directed graph defines the following class of probabilistic models

$$p(x_1, \ldots x_6) = p(x_1)p(x_2)p(x_3|x_1)p(x_4|x_1, x_2)p(x_5|x_4)p(x_6|x_4)$$

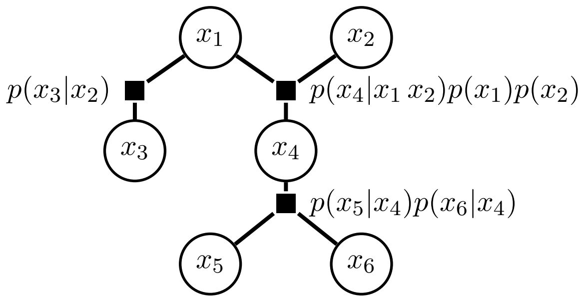
This gives the factor graph on right in the figure below.





Note:

- The moral graph contains a loop while the factor graph does not. The factor graph is still a polytree. This can be exploited for inference.
- One may choose to group some factors together in order to obtain a factor graph with a particular structure (see factor graph below)



Exercise 3. ***Limits of directed and undirected graphical models***

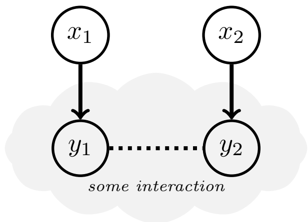
We here consider the probabilistic model $p(y_1, y_2, x_1, x_2) = p(y_1, y_2|x_1, x_2)p(x_1)p(x_2)$ where $p(y_1, y_2|x_1, x_2)$ factorises as

$$p(y_1, y_2|x_1, x_2) = p(y_1|x_1)p(y_2|x_2)\phi(y_1, y_2)n(x_1, x_2) \tag{1}$$

with $n(x_1, x_2)$ equal to

$$n(x_1, x_2) = \left(\int p(y_1|x_1)p(y_2|x_2)\phi(y_1, y_2)\mathrm{d}y_1\mathrm{d}y_2\right)^{-1}. \tag{2}$$

In the lecture, we used the model to illustrate the setup where x_1 and x_2 are two independent inputs that each control the interacting variables y_1 and y_2 (see graph below).



(a) Use the basic characterisations of statistical independence

$$u \perp\!\!\!\perp v|z \Longleftrightarrow p(u, v|z) = p(u|z)p(v|z) \tag{3}$$

$$u \perp\!\!\!\perp v|z \Longleftrightarrow p(u, v|z) = a(u, z)b(v, z) \qquad (a(u, z) \geq 0, b(v, z) \geq 0) \tag{4}$$

to show that $p(y_1, y_2, x_1, x_2)$ satisfies the following independencies

- $x_1 \perp\!\!\!\perp x_2$ *(independence between control variables)*
- $x_1 \perp\!\!\!\perp y_2 \mid y_1, x_2$ *(y_2 is only influenced by y_1 and x_2)*
- $x_2 \perp\!\!\!\perp y_1 \mid y_2, x_1$ *(y_1 is only influenced by y_2 and x_1)*

Solution. The pdf/pmf is

$$p(y_1, y_2, x_1, x_2) = p(y_1|x_1)p(y_2|x_2)\phi(y_1, y_2)n(x_1, x_2)p(x_1)p(x_2)$$

For $\mathbf{x_1} \perp\!\!\!\perp \mathbf{x_2}$
We compute $p(x_1, x_2)$ as

$$\begin{aligned}
p(x_1, x_2) &= \int p(y_1, y_2, x_1, x_2)\mathrm{d}y_1\mathrm{d}y_2 && \text{(S.1)}\\
&= \int p(y_1|x_1)p(y_2|x_2)\phi(y_1, y_2)n(x_1, x_2)p(x_1)p(x_2)\mathrm{d}y_1\mathrm{d}y_2 && \text{(S.2)}\\
&= n(x_1, x_2)p(x_1)p(x_2)\int p(y_1|x_1)p(y_2|x_2)\phi(y_1, y_2)\mathrm{d}y_1\mathrm{d}y_2 && \text{(S.3)}\\
&\overset{(2)}{=} n(x_1, x_2)p(x_1)p(x_2)\frac{1}{n(x_1, x_2)} && \text{(S.4)}\\
&= p(x_1)p(x_2). && \text{(S.5)}
\end{aligned}$$

Since $p(x_1)$ and $p(x_2)$ are the univariate marginals of x_1 and x_2, respectively, it follows from (3) that $x_1 \perp\!\!\!\perp x_2$.

For $\mathbf{x_1} \perp\!\!\!\perp \mathbf{y_2} \mid \mathbf{y_1}, \mathbf{x_2}$
We rewrite $p(y_1, y_2, x_1, x_2)$ as

$$\begin{aligned}
p(y_1, y_2, x_1, x_2) &= p(y_1|x_1)p(y_2|x_2)\phi(y_1, y_2)n(x_1, x_2)p(x_1)p(x_2) && \text{(S.6)}\\
&= [p(y_1|x_1)p(x_1)n(x_1, x_2)]\,[p(y_2|x_2)\phi(y_1, y_2)p(x_2)] && \text{(S.7)}\\
&= \phi_A(x_1, y_1, x_2)\phi_B(y_2, y_1, x_2) && \text{(S.8)}
\end{aligned}$$

With (4), we have that $x_1 \perp\!\!\!\perp y_2 \mid y_1, x_2$. Note that $p(x_2)$ can be associated either with ϕ_A or with ϕ_B.

For $\mathbf{x_2} \perp\!\!\!\perp \mathbf{y_1} \mid \mathbf{y_2}, \mathbf{x_1}$
We use here the same approach as for $x_1 \perp\!\!\!\perp y_2 \mid y_1, x_2$. (By symmetry considerations, we could immediately see that the relation holds but let us write it out for clarity). We rewrite $p(y_1, y_2, x_1, x_2)$ as

$$\begin{aligned}
p(y_1, y_2, x_1, x_2) &= p(y_1|x_1)p(y_2|x_2)\phi(y_1, y_2)n(x_1, x_2)p(x_1)p(x_2) && \text{(S.9)}\\
&= [p(y_2|x_2)n(x_1, x_2)p(x_2)p(x_1))]\,[p(y_1|x_1)\phi(y_1, y_2)]) && \text{(S.10)}\\
&= \tilde{\phi}_A(x_2, x_1, y_2)\tilde{\phi}_B(y_1, y_2, x_1) && \text{(S.11)}
\end{aligned}$$

With (4), we have that $x_2 \perp\!\!\!\perp y_1 \mid y_2, x_1$.

(b) *Draw the undirected graph for $p(y_1, y_2, x_1, x_2)$ and check whether graph separation allows us to see all independencies listed above.*

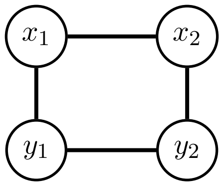
Solution. We write

$$p(y_1, y_2, x_1, x_2) = p(y_1|x_1)p(y_2|x_2)\phi(y_1, y_2)n(x_1, x_2)p(x_1)p(x_2)$$

as

$$\begin{aligned} p(y_1, y_2, x_1, x_2) &= \phi_1(y_1, x_1)\phi_2(y_2, x_2)\phi_3(y_1, y_2)\phi_4(x_1, x_2) \qquad \text{with} && \text{(S.12)} \\ \phi_1(y_1, x_1) &= p(y_1|x_1)p(x_1) && \text{(S.13)} \\ \phi_2(y_2, x_2) &= p(y_2|x_2)p(x_2) && \text{(S.14)} \\ \phi_3(y_1, y_2) &= \phi(y_1, y_2) && \text{(S.15)} \\ \phi_4(x_1, x_2) &= n(x_1, x_2) && \text{(S.16)} \end{aligned}$$

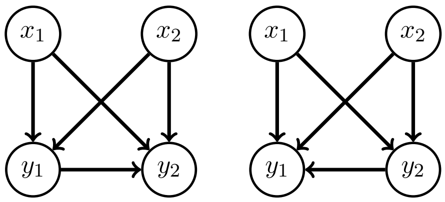
The corresponding undirected graph is as follows.



While the graph implies $x_1 \perp\!\!\!\perp y_2 \mid y_1, x_2$ and $x_2 \perp\!\!\!\perp y_1 \mid y_2, x_1$, the independency $x_1 \perp\!\!\!\perp x_2$ is not represented.

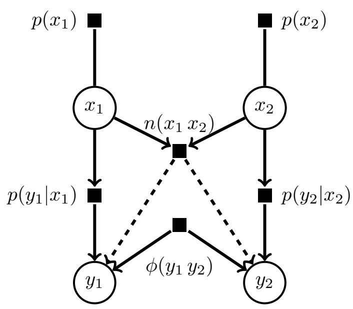
(c) *Draw the directed graph for $p(y_1, y_2, x_1, x_2) = p(y_1, y_2|x_1, x_2)p(x_1)p(x_2)$ and check whether graph separation allows us to see all independencies listed above.*

Solution. If we use the ordering x_1, x_2, y_1, y_2, we obtain the graph on the left. If we use the ordering x_1, x_2, y_2, y_1, we obtain the graph on the right.



The graphs do represent $x_1 \perp\!\!\!\perp x_2$ but not $x_1 \perp\!\!\!\perp y_2 \mid y_1, x_2$ and $x_2 \perp\!\!\!\perp y_1 \mid y_2, x_1$. Moreover, the graphs imply a directionality between y_1 and y_2, and a direct influence of x_1 on y_2, and of x_2 on y_1, in contrast to the original modelling goals.

(d) (optional, not examinable) *In the lecture, we have the following factor graph for $p(y_1, y_2, x_1, x_2)$*



Use the separation rules for factor graphs to verify that we can find all independence relations. The separation rules are (see Barber, section 4.4.1, or the original paper by Brendan Frey: https://arxiv.org/abs/1212.2486):

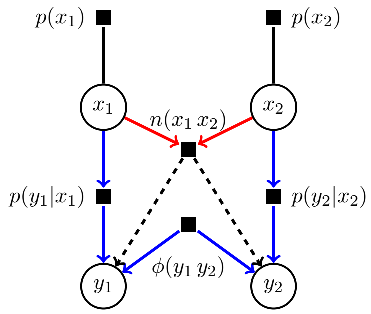
"If all paths are blocked, the variables are conditionally independent. A path is blocked if one or more of the following conditions is satisfied:

1. *One of the variables in the path is in the conditioning set.*
2. *One of the variables or factors in the path has two incoming edges that are part of the path (variable or factor collider), and neither the variable or factor nor any of its descendants are in the conditioning set."*

Remarks:

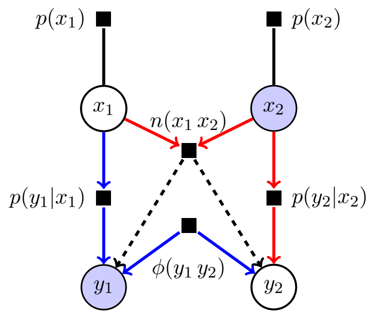
- *"one or more of the following" should best be read as "one of the following".*
- *"incoming edges" means directed incoming edges*
- *the descendants of a variable of factor node are all the variables that you can reach by following a path (containing directed or directed edges, but for directed edges, all directions have to be consistent)*
- *In the graph we have dashed directed edges: they do count when you determine the descendants but they do not contribute to paths. For example, y_1 is a descendant of the $n(x_1, x_2)$ factor node but $x_1 - n - y_2$ is not a path.*

Solution. $\mathbf{x_1 \perp\!\!\!\perp x_2}$
There are two paths from x_1 to x_2 marked with red and blue below:



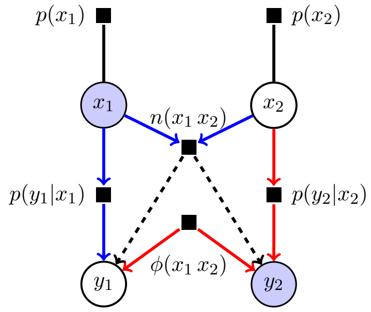
Both the blue and red path are blocked by condition 2.

$\mathbf{x_1 \perp\!\!\!\perp y_2 \mid y_1, x_2}$
There are two paths from x_1 to y_2 marked with red and blue below:



The observed variables are marked in blue. For the red path, the observed x_2 blocks the path (condition 1). Note that the $n(x_1, x_2)$ node would be open by condition 2. The blue path is blocked by condition 1 too. In directed graphical models, the y_1 node would be open, but here while condition 2 does not apply, condition 1 still applies (note the *one or more of ...* in the separation rules), so that the path is blocked.

$\mathbf{x_2} \perp\!\!\!\perp \mathbf{y_1} \mid \mathbf{y_2}, \mathbf{x_1}$
There are two paths from x_2 to y_1 marked with red and blue below:



The same reasoning as before yields the result.

Finally note that x_1 and x_2 are not independent given y_1 or y_2 because the upper path through $n(x_1, x_2)$ is not blocked whenever y_1 or y_2 are observed (condition 2).

Credit: this example is discussed in the original paper by B. Frey (Figure 6).