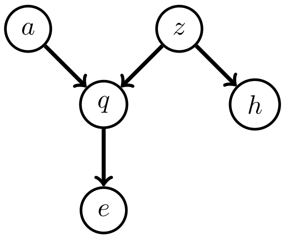
Exercise 1. ***Ordered and local Markov properties, d-separation***

Consider the graph below:



(a) *The ordering (z, h, a, q, e) is topological to the graph. What are the independencies that follow from the ordered Markov property?*

Solution. We proceed as in the lecture slides: The predecessor sets are

$$\mathrm{pre}_z = \varnothing, \mathrm{pre}_h = \{z\}, \mathrm{pre}_a = \{z, h\}, \mathrm{pre}_q = \{z, h, a\}, \mathrm{pre}_e = \{z, h, a, q\}$$

The parent sets are independent from the topological ordering chosen. In the lecture, we have seen that they are:

$$\mathrm{pa}_z = \varnothing, \mathrm{pa}_h = \{z\}, \mathrm{pa}_a = \varnothing, \mathrm{pa}_q = \{a, z\}, \mathrm{pa}_e = \{q\},$$

The ordered Markov property reads $x_i \perp\!\!\!\perp (\mathrm{pre}_i \setminus \mathrm{pa}_i) \mid \mathrm{pa}_i$ where the x_i refer to the ordered variables, e.g. $x_1 = z, x_2 = h, x_3 = a, etc.$
With

$$\mathrm{pre}_h \setminus \mathrm{pa}_h = \varnothing \quad \mathrm{pre}_a \setminus \mathrm{pa}_a = \{z, h\} \quad \mathrm{pre}_q \setminus \mathrm{pa}_q = \{h\} \quad \mathrm{pre}_e \setminus \mathrm{pa}_e = \{z, h, a\}$$

we thus obtain

$$h \perp\!\!\!\perp \varnothing \mid z \qquad a \perp\!\!\!\perp \{z, h\} \qquad q \perp\!\!\!\perp h \mid \{a, z\} \qquad e \perp\!\!\!\perp \{z, h, a\} \mid q$$

The relation $h \perp\!\!\!\perp \varnothing \mid z$ should be understood as "there is no variable from which h is independent given z" and should thus be dropped from the list. Compared to the relations obtained for the orderings in the lecture, the new one here is $a \perp\!\!\!\perp \{z, h\}$. Generally, having a variable later in the topological ordering allows one to possibly obtain a stronger independence relation because the set pre $\setminus$ pa can only increase when the predecessor set pre becomes larger.

(b) *What are the independencies that follow from the local Markov property?*

Solution. The non-descendants are

$$\mathrm{nondesc}(a) = \{z, h\} \quad \mathrm{nondesc}(z) = \{a\} \quad \mathrm{nondesc}(h) = \{a, z, q, e\}$$
$$\mathrm{nondesc}(q) = \{a, z, h\} \quad \mathrm{nondesc}(e) = \{a, q, z, h\}$$

With the parent sets as before, the independencies that follow from the local Markov property are $x_i \perp\!\!\!\perp (\mathrm{nondesc}(x_i) \setminus \mathrm{pa}_i) \mid \mathrm{pa}_i$, i.e.

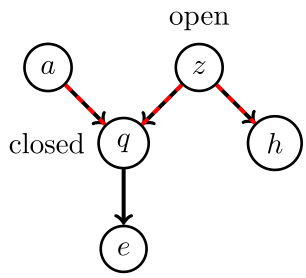
$$a \perp\!\!\!\perp \{z, h\} \qquad z \perp\!\!\!\perp a \qquad h \perp\!\!\!\perp \{a, q, e\} \mid z \qquad q \perp\!\!\!\perp h \mid \{a, z\} \qquad e \perp\!\!\!\perp \{a, z, h\} \mid q$$

(c) The independency relations obtained via the ordered and local Markov property include $q \perp\!\!\!\perp h \mid \{a, z\}$. Verify the independency using d-separation.

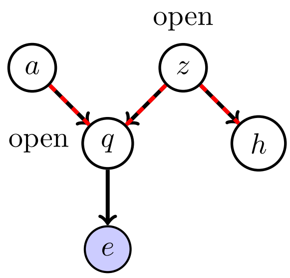
Solution. The only trail from q to h goes through z which is in a tail-tail configuration. Since z is part of the conditioning set, the trail is blocked and the result follows.

(d) Use d-separation to check whether $a \perp\!\!\!\perp h \mid e$ holds.

Solution. The trail from a to h is shown below in red together with the default states of the nodes along the trail.



Conditioning on e opens the q node since q in a collider configuration on the path.



The trail from a to h is thus active, which means that the relationship does not hold because $a \not\perp\!\!\!\perp h \mid e$ for some distributions that factorise over the graph.

(e) Assume all variables in the graph are binary. How many numbers do you need to specify, or learn from data, in order to fully specify the probability distribution?

Solution. The graph defines a set of probability mass functions (pmf) that factorise as

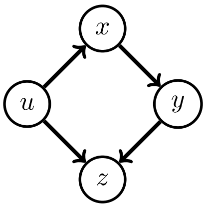
$$p(a, z, q, h, e) = p(a)p(z)p(q|a, z)p(h|z)p(e|q)$$

To specify a member of the set, we need to specify the (conditional) pmfs on the right-hand side. The (conditional) pmfs can be seen as tables, and the number of elements that we need to specified in the tables are:
- 1 for $p(a)$
- 1 for $p(z)$
- 4 for $p(q|a, z)$
- 2 for $p(h|z)$
- 2 for $p(e|q)$

In total, there are 10 numbers to specify. This is in contrast to $2^5 - 1 = 31$ for a distribution without independencies. Note that the number of parameters to specify could be further reduced by making parametric assumptions.

Exercise 2. *Flipping arrows*

Consider the following graph:



(a) *How does the corresponding directed graphical model factorise?*

Solution. $p(u, x, y, z) = p(u)p(x|u)p(y|x)p(z|y, u)$

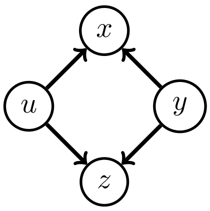
(b) *List all independencies encoded by the graph.*

Solution. Variable u cannot be independent of x and z since they are connected. Since z is in a collider configuration, we have that $u \perp\!\!\!\perp y|x$.

Variable x cannot be independent of y (and u) since they are connected. Conditioning on u and y makes x independent of z: $x \perp\!\!\!\perp z|u, y$

(c) *How do the independencies change when we flip the arrow from x to y so that it points the other way around, i.e. $y \to x$.?*

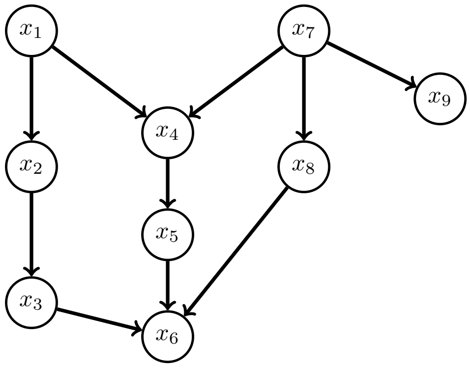
Solution. With the change, we obtain the following graph



which has two colliders blocking all paths between u and y. Hence $u \perp\!\!\!\perp y$. However, conditioning on x would open the path so that $u \not\perp\!\!\!\perp y|x$. The independency $x \perp\!\!\!\perp z|u, y$ still holds.

Exercise 3. *Independencies in directed graphical models*

Consider the following directed acyclic graph.



For each of the statements below, determine whether it holds for all probabilistic models that factorise over the graph. Provide a justification for your answer.

(a) $p(x_7|x_2) = p(x_7)$

Solution. Yes, it holds. x_2 is a non-descendant of x_7, $\text{pa}(x_7) = \varnothing$, and hence, by the local Markov property, $x_7 \perp\!\!\!\perp x_2$, so that $p(x_7|x_2) = p(x_7)$.

(b) $x_1 \not\perp\!\!\!\perp x_3$

Solution. No, does not hold. x_1 and x_3 are d-connected, which only implies independence for *some* and not all distributions that factorise over the graph. The graph generally only allows us to read out independencies and not dependencies.

(c) $p(x_1, x_2, x_4) \propto \phi_1(x_1, x_2)\phi_2(x_1, x_4)$ *for some non-negative functions* ϕ_1 *and* ϕ_2.

Solution. Yes, it holds. The statement is equivalent to $x_2 \perp\!\!\!\perp x_4 \mid x_1$. There are three trails from x_2 to x_4, which are all blocked:

1. $x_2 - x_1 - x_4$: this trail is blocked because x_1 is in a tail-tail connection and it is observed, which closes the node.
2. $x_2 - x_3 - x_6 - x_5 - x_4$: this trail is blocked because x_3, x_6, x_5 is in a collider configuration, and x_6 is not observed (and it does not have any descendants).
3. $x_2 - x_3 - x_6 - x_8 - x_7 - x_4$: this trail is blocked because x_3, x_6, x_8 is in a collider configuration, and x_6 is not observed (and it does not have any descendants).

Hence, by the global Markov property (d-separation), the independency holds.

(d) $x_2 \perp\!\!\!\perp x_9 \mid \{x_6, x_8\}$

Solution. No, does not hold. Conditioning on x_6 opens the collider node x_4 on the trail $x_2 - x_1 - x_4 - x_7 - x_9$, so that the trail is active.

(e) $x_8 \perp\!\!\!\perp \{x_2, x_9\} \mid \{x_3, x_5, x_6, x_7\}$

Solution. Yes, it holds. $\{x_3, x_5, x_6, x_7\}$ is the Markov blanket of x_8, so that x_8 is independent of remaining nodes given the Markov blanket.

(f) $\mathbb{E}[x_2 \cdot x_3 \cdot x_4 \cdot x_5 \cdot x_8 \mid x_7] = 0$ *if* $\mathbb{E}[x_8 \mid x_7] = 0$

Solution. Yes, it holds. $\{x_2, x_3, x_4, x_5\}$ are non-descendants of x_8, and x_7 is the parent of x_8, so that $x_8 \perp\!\!\!\perp \{x_2, x_3, x_4, x_5\} \mid x_7$. This means that $\mathbb{E}[x_2 \cdot x_3 \cdot x_4 \cdot x_5 \cdot x_8 \mid x_7] = \mathbb{E}[x_2 \cdot x_3 \cdot x_4 \cdot x_5 \mid x_7]\mathbb{E}[x_8 \mid x_7] = 0$.