(a) p factorizes according to $\mathcal{G}$;

(b) p satisfies the directed global Markov property with respect to $\mathcal{G}$;

(c) p satisfies the directed local Markov property with respect to $\mathcal{G}$.

A useful intuition for the implication of (a)$\Rightarrow$ (b) in Theorem 1 comes from our three-node examples depicted in Fig. 3. In particular, we can think of paths through a middle node being "blocked" or not depending on both the direction of the edges incident to it, and whether that node is in the conditioning set or not.

More specifically, we saw that when we have the edge structure of Example 1, the variables associated with the end nodes are independent (corresponding to a blocked path) if conditioned on that associated with the middle node, and not otherwise. Moreover, the same holds for the common-cause structure of Example 2. However, the opposite holds for the common-effect (V-structure) of Example 3: the variables associated with the end nodes are independent without any conditioning on the remaining variable, and are otherwise dependent.

Checking for d-separation is straightforward to carry out on graphs of modest size, by inspection. For larger graphs, an efficient algorithm can be designed based on checking d-separation by inspecting all paths in breadth-first search manner. In the literature, this algorithm is known as the *Bayes ball* algorithm.

We now prove Theorem 1. We divide the proof into three parts: (1) if p factorizes according to $\mathcal{G}$, then d-separation implies conditional independence i.e., $(a) \Rightarrow (b)$, (2) if p satisfies the directed global Markov property with respect to $\mathcal{G}$, then it also satisfies the directed local Markov property with respect to $\mathcal{G}$ i.e., $(b) \Rightarrow (c)$, and (3) if p satisfies the directed local Markov property with respect to $\mathcal{G}$, then it factors according to $\mathcal{G}$ i.e., $(c) \Rightarrow (a)$. The proof of (1) is somewhat more involved while the others are pretty straightforward.

Proof. We start with (1), showing that d-separation implies conditional independence. We use induction on the number of nodes N. For $N = 1$, there is nothing to show. Suppose the statement holds for any directed acyclic graph (DAG) on $N - 1$ nodes. Now consider a DAG $\mathcal{D}$ on N nodes. Consider a topological ordering of all the nodes of $\mathcal{D}$ and let ω be the node with the largest number, i.e. it has no descendent or child. Without loss of generality relabel the nodes so that $\omega = N$. Let $\mathcal{D}'$ be obtained by removing ω from $\mathcal{D}$. It is also a DAG and has $N - 1$ nodes. Crucially, the set of nodes $1, \ldots, N-1$ have a distribution that factorizes according to the DAG $\mathcal{D}'$: due to the factorization implied by $\mathcal{D}$, we can write out the joint distribution $p_{\mathsf{x}_1,\ldots,\mathsf{x}_{N-1}}$ as the product $\prod_{i=1}^{N-1} p_{\mathsf{x}_i|\mathsf{x}_{\pi_i}}(x_i|x_{\pi_i})$, and this is exactly the factorization required by $\mathcal{D}'$. Note that this is a consequence of the choice of ω as the last node in the topological ordering. There are three possibilities stated below and for each of them we prove the desired statement using the induction hypothesis:

(a) $\omega \notin \mathcal{A} \cup \mathcal{B} \cup \mathcal{C}$. Consider a pair of nodes $a \in \mathcal{A}$ and $b \in \mathcal{B}$ and any path between them in $\mathcal{D}'$. This path can be viewed as a path in $\mathcal{D}$, and we have by assumption that it is blocked in $\mathcal{D}$ by $\mathcal{C}$. If this blocking is via rule C1 (above Defn 2), then some node $c \in \mathcal{C}$, blocks the path in $\mathcal{D}$ and continues to do so in $\mathcal{D}'$. If the blocking is via rule C2, i.e., there is some node on the path where the arrows meet head-to-head and neither it nor any of its descendants $\mathcal{D}$ are in $\mathcal{C}$, we again see that this continues to hold in $\mathcal{D}'$. Thus the path is blocked in $\mathcal{D}'$ and by the induction hypothesis applied to $\mathcal{D}'$, we obtain that $\mathsf{x}_{\mathcal{A}} \perp\!\!\!\perp \mathsf{x}_{\mathcal{B}} \mid \mathsf{x}_{\mathcal{C}}$.

(b) $\omega \in \mathcal{A}$ (the same argument applies to $\omega \in \mathcal{B}$). Let $\mathcal{A}' = \mathcal{A}\backslash\omega$. Note that $\omega \notin \mathcal{A}' \cup \mathcal{B} \cup \mathcal{C}$, so the previous paragraph shows that $\mathcal{A}'$ is d-separated from $\mathcal{B}$ with respect to $\mathcal{C}$ in $\mathcal{D}'$ and therefore $\mathsf{x}_{\mathcal{A}'} \perp\!\!\!\perp \mathsf{x}_{\mathcal{B}} \mid \mathsf{x}_{\mathcal{C}}$.

Observe that no parent of ω is in set $\mathcal{B}$: Otherwise, we have an edge between ω and a node in $\mathcal{B}$, which violates the assumption of d-separation of $\mathcal{A}$ and $\mathcal{B}$ given $\mathcal{C}$. Let $P = \pi_\omega \setminus \mathcal{C}$ be the set of parent nodes of ω in $\mathcal{D}$ that are not in $\mathcal{C}$. We will show that (i) $\mathsf{x}_{\mathcal{A}' \cup P} \perp\!\!\!\perp \mathsf{x}_{\mathcal{B}} \mid \mathsf{x}_{\mathcal{C}}$, and (ii) $\mathsf{x}_\omega \perp\!\!\!\perp \mathsf{x}_{\mathcal{B}} \mid \mathsf{x}_{\mathcal{A}' \cup \mathcal{C} \cup P}$. From these and the definition of conditional independence one can obtain $\mathsf{x}_{\mathcal{A}' \cup P \cup \omega} \perp\!\!\!\perp \mathsf{x}_{\mathcal{B}} \mid \mathsf{x}_{\mathcal{C}}$ (be sure to check this for yourself!).

To show (i), it is sufficient to show that $\mathsf{x}_P \perp\!\!\!\perp \mathsf{x}_{\mathcal{B}} \mid \mathsf{x}_{\mathcal{C}}$ and then to combine this with $\mathsf{x}_{\mathcal{A}'} \perp\!\!\!\perp \mathsf{x}_{\mathcal{B}} \mid \mathsf{x}_{\mathcal{C} \cup P}$. We start by showing the latter conditional independence, which is similar to $\mathsf{x}_{\mathcal{A}'} \perp\!\!\!\perp \mathsf{x}_{\mathcal{B}} \mid \mathsf{x}_{\mathcal{C}}$ shown in case (a) above, and but conditioning on $\mathsf{x}_{\mathcal{C} \cup P}$ instead of $\mathsf{x}_{\mathcal{C}}$. Consider a path from $a \in \mathcal{A}'$ to $b \in \mathcal{B}$ in $\mathcal{D}'$. The same argument as in (a) shows that if it is blocked via rule C1 by $\mathcal{C}$ in $\mathcal{D}$, then it remains so in $\mathcal{D}'$ and also when adding in P to the conditioning set, and the desired statement follows by the inductive hypothesis on $\mathcal{D}'$ as before. The remaining case to consider is that the blocking of the path under consideration is (only) via rule C2, in which case adding nodes P to the conditioning set can plausibly unblock the path between a and b. Suppose for the sake of contradiction that this occurs. This can happen only if a head-to-head node in the path has a node in P as its descendant, and let d be the node that is closest to b among such nodes. Since ω is a child of p in $\mathcal{D}$, we see that there is a path between $b \in B$ and $\omega \in A$ in $\mathcal{D}$ that is unblocked by $\mathcal{C}$, which contradicts our initial assumption. It follows that $\mathsf{x}_{\mathcal{A}'} \perp\!\!\!\perp \mathsf{x}_{\mathcal{B}} \mid \mathsf{x}_{\mathcal{C} \cup P}$.

To show $\mathsf{x}_P \perp\!\!\!\perp \mathsf{x}_{\mathcal{B}} \mid \mathsf{x}_{\mathcal{C}}$, we will argue that P is is d-separated from $\mathcal{B}$ by $\mathcal{C}$ in $\mathcal{D}'$ and use the induction hypothesis for $\mathcal{D}'$. Suppose for the sake of contradiction that this were not the case, i.e., there is a path in $\mathcal{D}'$ from a node $p \in P$ to a node $b \in \mathcal{B}$ that it is not blocked with respect to $\mathcal{C}$. Now extend this path by adding ω as the first node (there is an edge between ω and p, oriented towards ω). Node p has become an internal node on the path. It is not head-to-head since the edge to ω is oriented towards ω, and it does not belong to $\mathcal{C}$ because we are in the case that $\omega \in \mathcal{A}$. We have constructed an unblocked path in $\mathcal{D}$ from $\omega \in \mathcal{A}$ to $b \in \mathcal{B}$ with respect to $\mathcal{C}$, which violates our assumption of d-separation. Therefore, P is d-separated from $\mathcal{B}$ with respect to $\mathcal{C}$ in $\mathcal{D}'$, which proves $\mathsf{x}_P \perp\!\!\!\perp \mathsf{x}_{\mathcal{B}} \mid \mathsf{x}_{\mathcal{C}}$ using the induction hypothesis.

Now we move on to (ii) $\mathsf{x}_\omega \perp\!\!\!\perp \mathsf{x}_{\mathcal{B}} \mid \mathsf{x}_{\mathcal{A}' \cup \mathcal{C} \cup P}$. For this, again note that all parents of ω are a subset of $\mathcal{A} \cup \mathcal{C} \cup P$ and $\mathcal{B}$ are ancestors disjoints from these. Therefore, by the Markov property of the DAG we obtain (ii).

(c) Finally, consider the case $\omega \in \mathcal{C}$. Let $\mathcal{C}' = \mathcal{C}\backslash\omega$. Then, we claim that $\mathcal{A}$ and $\mathcal{B}$ must be d-separated by $\mathcal{C}'$ in $\mathcal{D}'$. Consider an arbitrary path between nodes $a \in \mathcal{A}$ and $b \in \mathcal{B}$, and by assumption it is blocked by $\mathcal{C}$ in $\mathcal{D}$. If the blockage is via rule C1, then it cannot be ω that is responsible, since ω has no descendants and both of the relevant scenarios require the blocking node to have an outgoing edge. If the blockage is via rule C2, then removing ω does not change presence of a head-to-head node with no ancestors in $\mathcal{C}'$ as $\mathcal{C}' \subset \mathcal{C}$. Thus, $\mathcal{A}$ and $\mathcal{B}$ are d-separated by $\mathcal{C}'$ in $\mathcal{D}'$, and by the induction hypothesis we have that $\mathsf{x}_{\mathcal{A}} \perp\!\!\!\perp \mathsf{x}_{\mathcal{B}} \mid \mathsf{x}_{\mathcal{C}'}$.

Now ω must be d-separated from at least one of $\mathcal{A}$ or $\mathcal{B}$ with respect to $\mathcal{C}'$. Suppose not. That is, ω has unblocked path to a node $a \in \mathcal{A}$ *and* has an unblocked path to a node $b \in \mathcal{B}$ with respect to $\mathcal{C}'$. Then, concatenation of these two paths with ω yields an unblocked path (since the two paths meet head-to-head at ω and hence rule (a) does not apply) between a and b with respect to $\mathcal{C}$ in $\mathcal{D}$. This contradicts our initial assumption and hence ω is d-separated either from $\mathcal{A}$ or $\mathcal{B}$ with respect to $\mathcal{C}'$. Without loss of generality, let ω be d-separated from $\mathcal{B}$ with respect to $\mathcal{C}'$. That is, $\mathcal{A} \cup \omega$ is d-separated from $\mathcal{B}$ with respect to $\mathcal{C}'$ in $\mathcal{D}$. We want to conclude that $\mathsf{x}_{\mathcal{A} \cup \omega} \perp\!\!\!\perp \mathsf{x}_{\mathcal{B}} \mid \mathsf{x}_{\mathcal{C}'}$.

To that end, we argue similarly to case (b). Consider parents P of ω that are not in $\mathcal{C}'$. As before, we can argue that P is d-separated from $\mathcal{B}$ with respect to $\mathcal{C}'$ in $\mathcal{D}'$. Therefore, $\mathsf{x}_{\mathcal{A} \cup P} \perp\!\!\!\perp \mathsf{x}_{\mathcal{B}} | \mathsf{x}_{\mathcal{C}'}$ using the induction hypothesis. Further, as argued earlier, it must be that $P \cap \mathcal{B} = \varnothing$. Therefore, using the property of DAG, we have that $\mathsf{x}_\omega \perp\!\!\!\perp \mathsf{x}_{\mathcal{B}} | \mathsf{x}_{\mathcal{A} \cup P \cup \mathcal{C}'}$ since $\mathcal{B}$ are ancestors of ω excluding it's parents and $\mathcal{A} \cup \mathcal{P} \cup \mathcal{C}'$ contains all parents of ω. Putting everything together, we have $\mathsf{x}_{\mathcal{A} \cup \omega} \perp\!\!\!\perp \mathsf{x}_{\mathcal{B}} \mid \mathsf{x}_{\mathcal{C}'}$. From this, it can be checked from definition of conditional independence that $\mathsf{x}_{\mathcal{A}} \perp\!\!\!\perp \mathsf{x}_{\mathcal{B}} \mid \mathsf{x}_{\mathcal{C}' \cup \omega}$. That is, $\mathsf{x}_{\mathcal{A}} \perp\!\!\!\perp \mathsf{x}_{\mathcal{B}} \mid \mathsf{x}_{\mathcal{C}}$.

We now move on to part (2), showing that the directed global Markov property implies the directed local Markov property. For any $i \in \mathcal{V}$, note that $\{i\}$ is d-separated from the set $nd(i) \setminus \pi_i$ with respect to π_i. Hence by the global Markov property, it holds that $\mathsf{x}_i \perp\!\!\!\perp \mathsf{x}_{nd(i)\setminus\pi_i} \mid \mathsf{x}_{\pi_i}$, which implies the local Markov property.

We now move on to part (3), showing that the directed local Markov property implies factorization according to $\mathcal{G}$. Without loss of generality assume a topological ordering of the nodes. Then using the chain rule, we can write th joint distribution as

$$p_{\mathsf{x}_1,\ldots,\mathsf{x}_N}(x_1, \ldots, x_N) = \prod_{i=1}^{N} p_{\mathsf{x}_i|\mathsf{x}_1,\ldots,\mathsf{x}_{i-1}}(x_i|x_1, \ldots, x_{i-1}) \tag{6}$$

$$= \prod_{i=1}^{N} p_{\mathsf{x}_i|\mathsf{x}_{\pi_i}}(x_i|x_{\pi_i}) \tag{7}$$

where the last equality follows from the directed local Markov property. Thus, we have that p factorizes according to the DAG $\mathcal{G}$, which concludes the proof. □