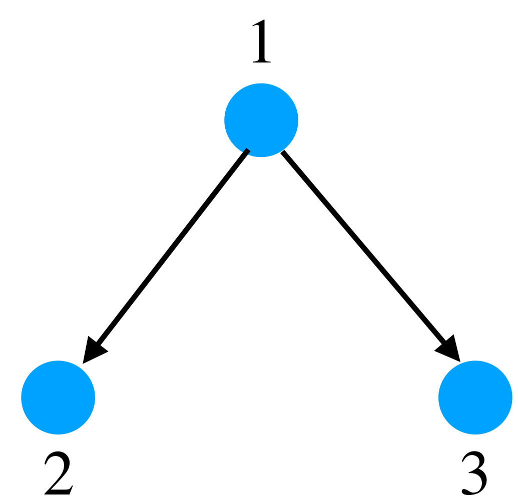
Graphical Models

Sullivant – Algebraic Statistics – Ch. 13

Roadmap



	Directed	Undirected
Multivariate Gaussian	?	?
Discrete	?	?

Graph + random variable indexed by the nodes

Implicit conditions

$$X_2 \perp\!\!\!\perp X_3 | X_1$$

Explicit parametrization

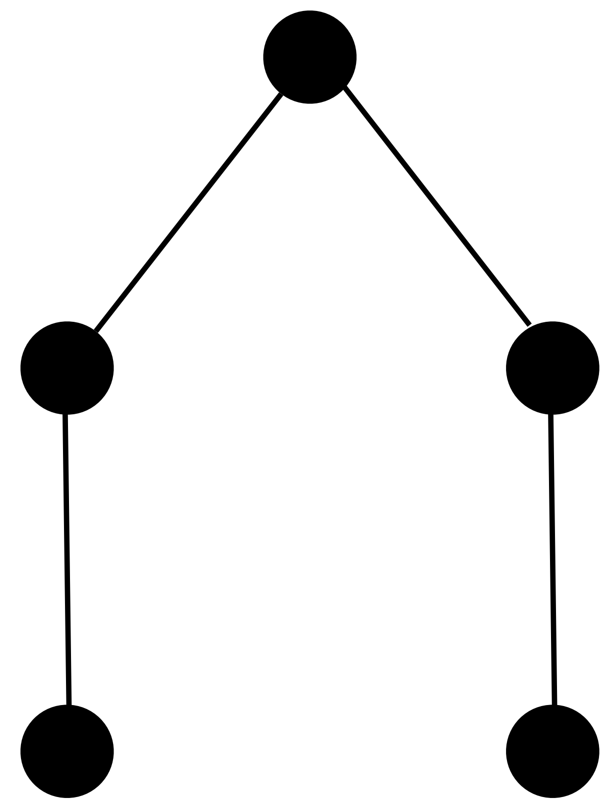
$$p(x) = p(x_1)p(x_2|x_1)p(x_3|x_1)$$

Undirected Graphs

Notation: $G = (V, E)$, V = vertices, E = edges.

$N(v)$ = neighbors of v.

X: random vector indexed by V.



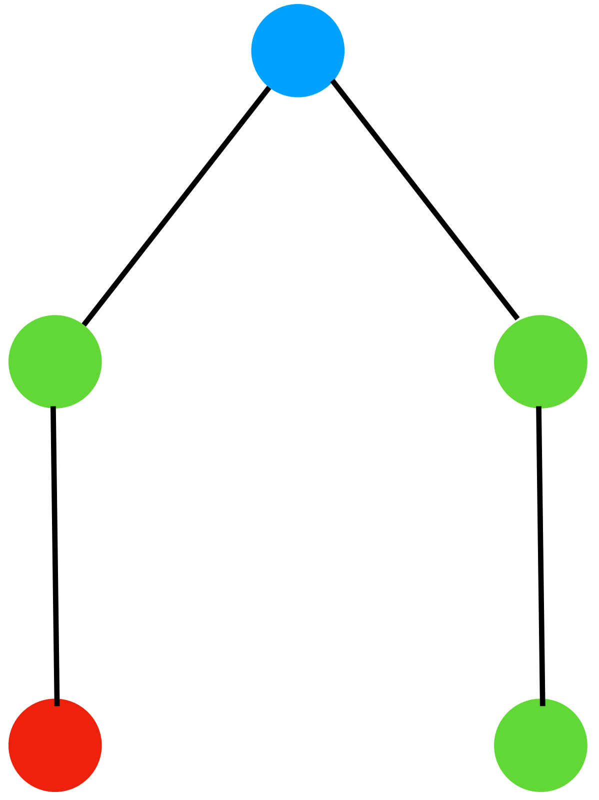
Markov Properties

Pairwise Markov property:

$$X_u \perp\!\!\!\perp X_v \mid X_{V \setminus \{u,v\}}$$

for all

$$\{u, v\} \notin E$$



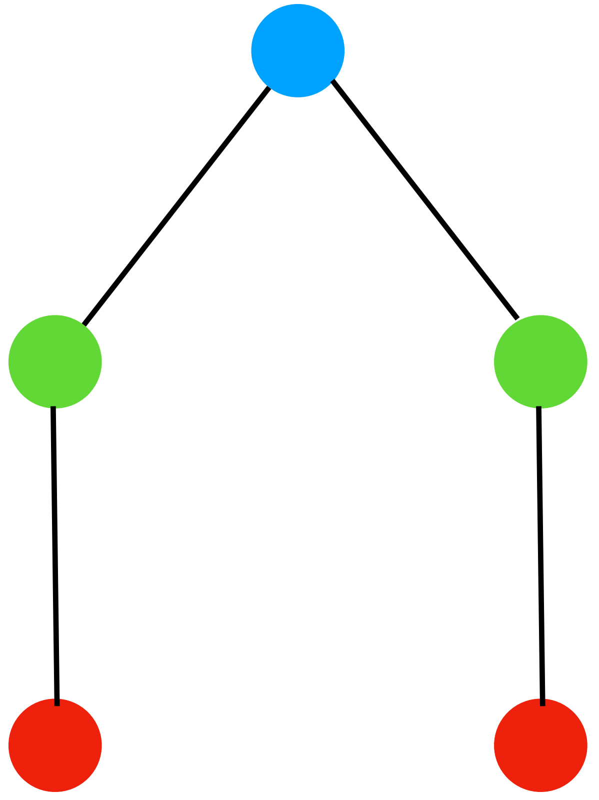
Markov Properties

Local Markov property

$$X_v \perp\!\!\!\perp X_{V \setminus (N(v) \cup v)} \mid X_{N(v)}$$

for all

$$v \in V$$

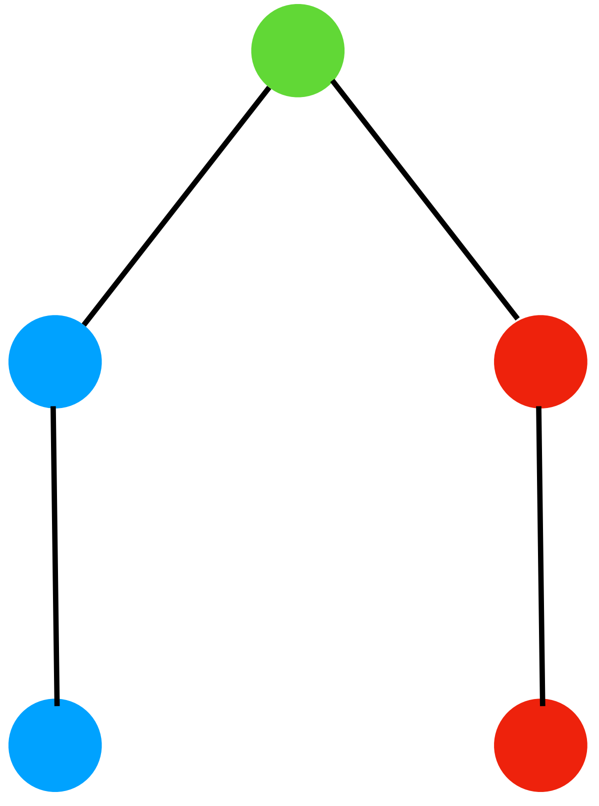


Markov Properties

Global Markov property

$$X_A \perp\!\!\!\perp X_B \mid X_C$$

for all disjoint A, B, C such that C separates A and B



Theorem 13.1.4

Intersection axiom ⇒ the Markov properties are equivalent.

In part.: if

$$P_X(x) > 0$$

for all x, then the Markov properties are equivalent for all G.

Example: Multivariate Gaussian

X: multivariate Gaussian with nonsingular covariance matrix Σ

Satisfies the intersection axiom

Example: Multivariate Gaussian

$$
\begin{aligned}
X_u \perp\!\!\!\perp X_v \mid X_{V\setminus\{u,v\}} &\Leftrightarrow \det \Sigma_{V\setminus u, V\setminus v} = 0 \\
&\Leftrightarrow \Sigma^{-1}_{u,v} = 0
\end{aligned}
$$

Example: Multivariate Gaussian

A, B, C disjoint subsets of V such that C does *not* separate A and B

$\Rightarrow$ there exists Σ such that X satisfies all global Markov statements but

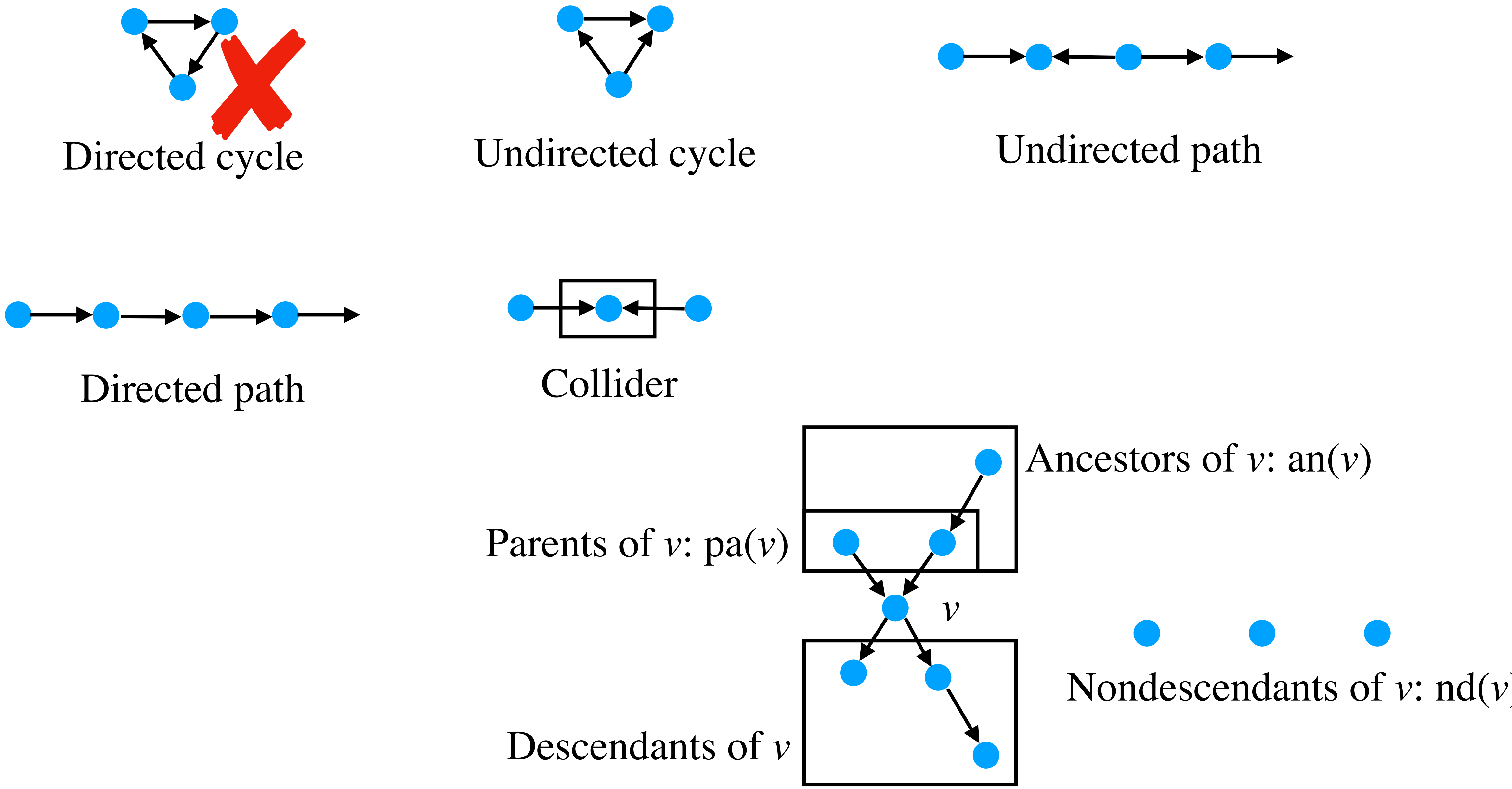
$$X_A \not\perp X_B \mid X_C$$

Path not sep. by C

$$\Sigma^{-1} = \begin{pmatrix} 1 & \varepsilon & & & & & \\ \varepsilon & 1 & \ddots & & & & \\ & \ddots & \ddots & \varepsilon & & & \\ & & \varepsilon & 1 & & & \\ & & & & 1 & & \\ & & & & & \ddots & \\ & & & & & & 1 \end{pmatrix}$$

Rest of the graph

Directed Acyclic Graphs



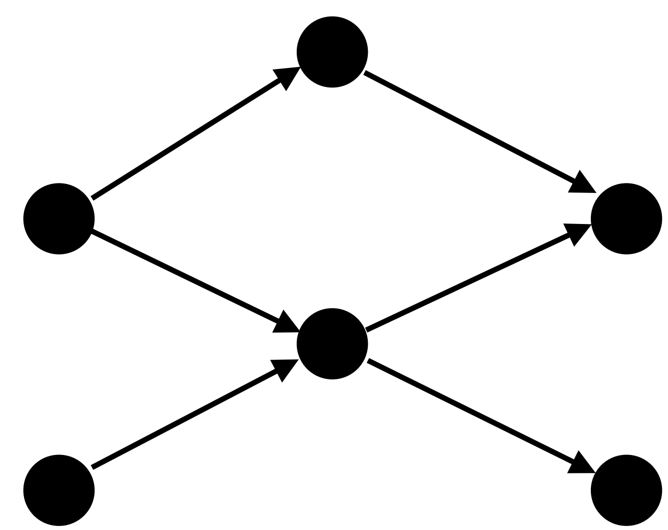
d-separation

$C \subset V,$ d-separates $v, w \in V:$

for all undirected paths π between v and w, in induced subgraph of π there exists

a collider in $C \cup \text{an}(C)$,

or a non-collider in C,



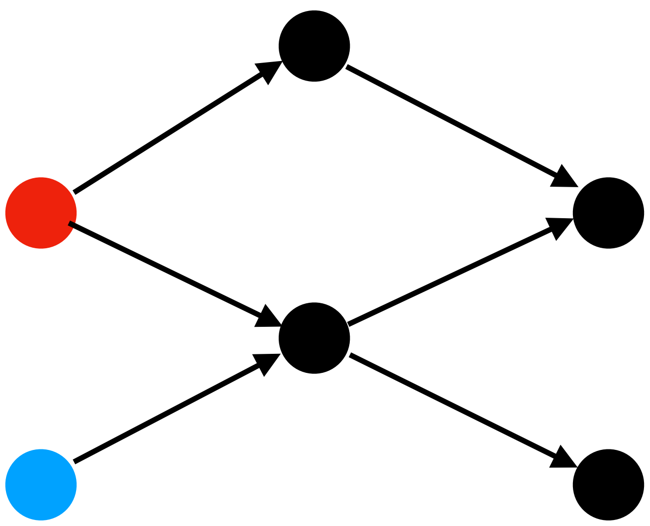
d-separation

$C \subset V$, d-separates $v, w \in V$:

for all undirected paths π between v and w, in induced subgraph of π there exists

a collider in $C \cup \mathrm{an}(C)$,

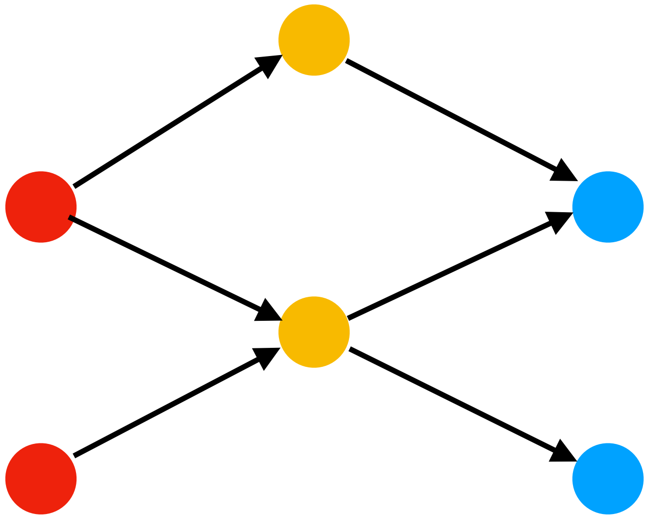
or a non-collider in C,



d-separation

$C \subset V$ d-separates $A, B \subset V$:

C d-separates all pairs $a \in A,\ b \in B$



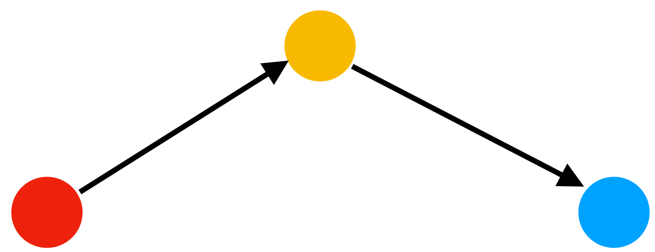
d-separation

$C \subset V$, d-separates $v, w \in V$:

for all undirected paths π between v and w, in induced subgraph of π there exists

a collider in $C \cup \mathrm{an}(C)$,

or a non-collider in C,



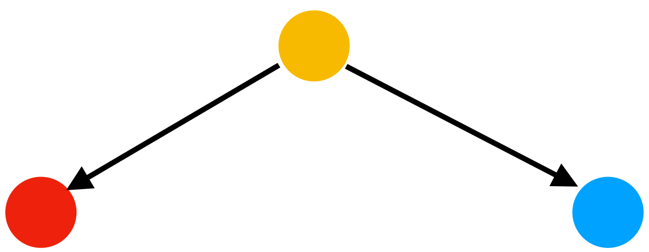
d-separation

$C \subset V$, d-separates $v, w \in V$:

for all undirected paths π between v and w, in induced subgraph of π there exists

a collider in $C \cup \text{an}(C)$,

or a non-collider in C,



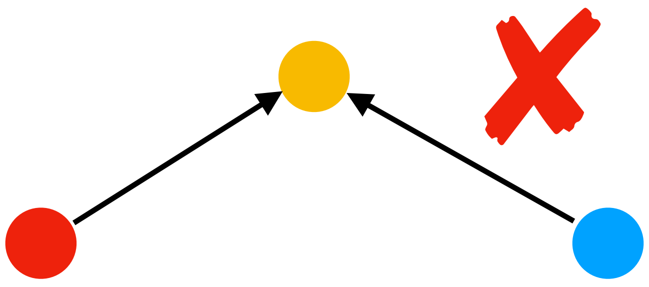
d-separation

$C \subset V,$ d-separates $v, w \in V:$

for all undirected paths π between v and w, in induced subgraph of π there exists

a collider in $C \cup \mathrm{an}(C)$,

or a non-collider in C,



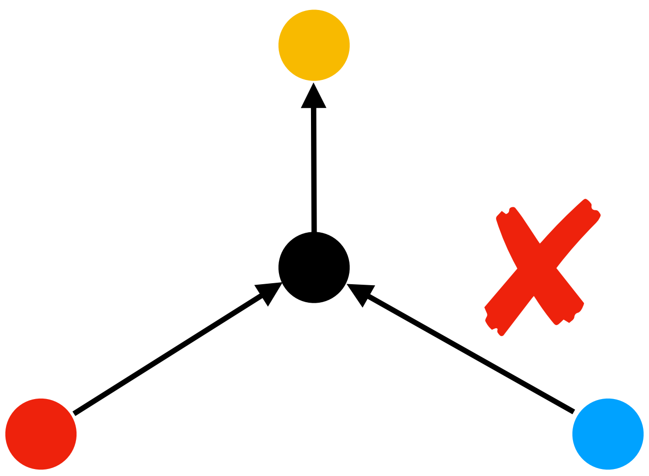
d-separation

$C \subset V,$ d-separates $v, w \in V:$

for all undirected paths π between v and w, in induced subgraph of π there exists

a collider in $C \cup \mathrm{an}(C)$,

or a non-collider in C,



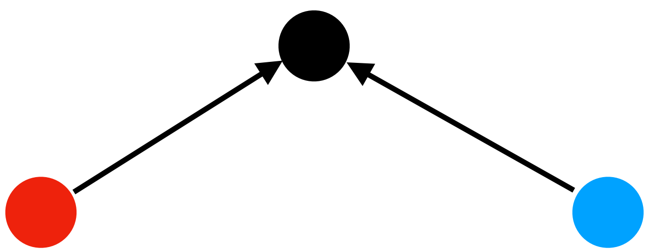
d-separation

$C \subset V$, d-separates $v, w \in V$:

for all undirected paths π between v and w, in induced subgraph of π there exists

a collider in $C \cup \mathrm{an}(C)$,

or a non-collider in C,



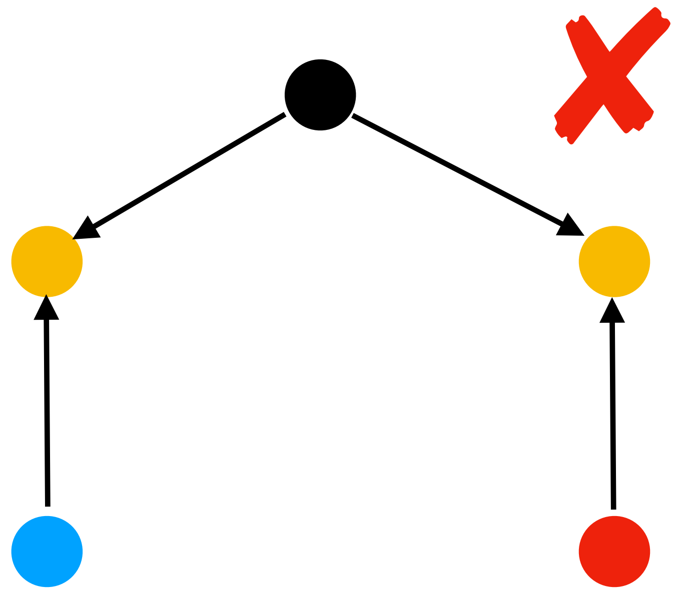
d-separation

$C \subset V$, d-separates $v, w \in V$:

for all undirected paths π between v and w, in induced subgraph of π there exists

a collider in $C \cup \mathrm{an}(C)$,

or a non-collider in C,



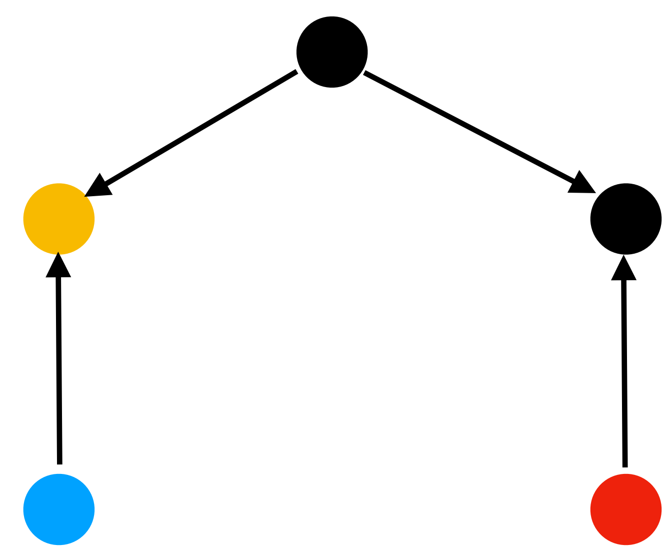
d-separation

$C \subset V$, d-separates $v, w \in V$:

for all undirected paths π between v and w, in induced subgraph of π there exists

a collider in $C \cup \mathrm{an}(C)$,

or a non-collider in C,



Markov Properties

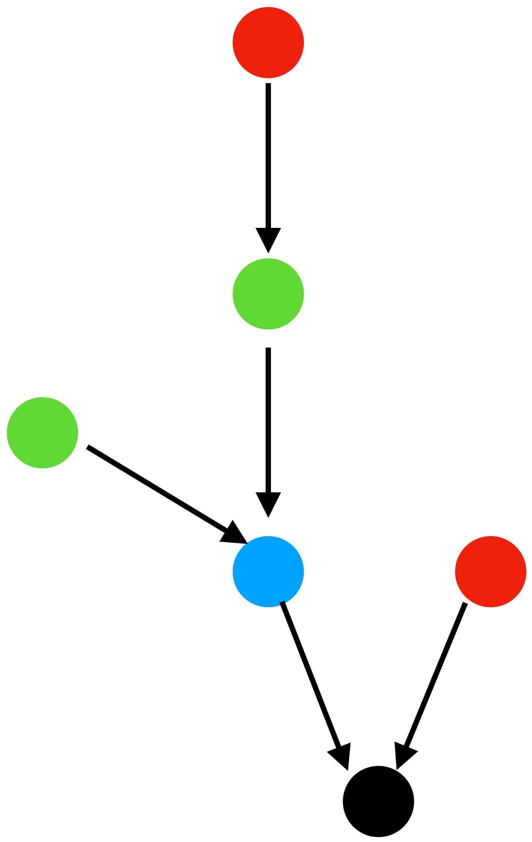
Definition 13.1.9. Let $G = (V, E)$ be a directed acyclic graph.

(i) The *directed pairwise Markov property* associated to G consists of all conditional independence statements $X_u \perp\!\!\!\perp X_v | X_{\mathrm{nd}(u) \setminus \{v\}}$ where (u, v) is not an edge of G.

(ii) The *directed local Markov property* associated to G consists of all conditional independence statements $X_v \perp\!\!\!\perp X_{\mathrm{nd}(v) \setminus \mathrm{pa}(v)} | X_{\mathrm{pa}(v)}$ for all $v \in V$.

(iii) The *directed global Markov property* associated to G consists of all conditional independence statements $X_A \perp\!\!\!\perp X_B | X_C$ for all disjoint sets A, B, and C such that C d-separates A and B in G.

Local Markov property



Parametrized undirected graphical model

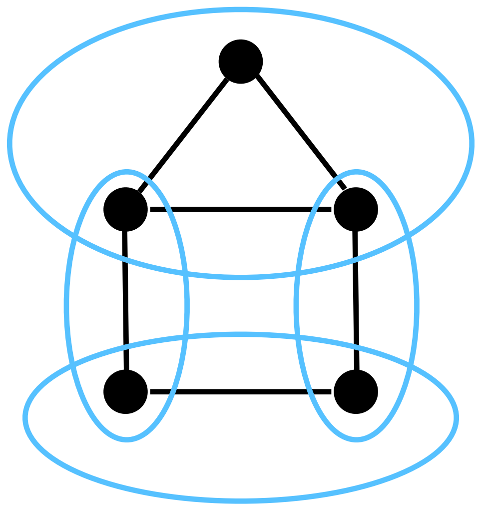
Density function:

$$f(x) = \frac{1}{Z} \prod_{C \in \mathcal{C}(G)} \phi_C(x_C)$$ "Factorizes according to G"

$\mathcal{C}(G)$**: set of all maximal cliques**

$\phi_C : X_C \to \mathbb{R}_{>0}$ **potential functions**

Z: **normalizing constant**



$$f(x_1, x_2, x_3, x_4, x_5) = \frac{1}{Z}\phi_{123}(x_1, x_2, x_3)\phi_{25}(x_2, x_5)\phi_{34}(x_3, x_4)\phi_{45}(x_4, x_5).$$

Theorem 13.2.3 (Hammersley–Clifford). *A continuous positive probability density f on $\mathcal{X}$ satisfies the pairwise Markov property on the graph G if and only if it factorizes according to G.*

Proof. $\Leftarrow$: let $f = \frac{1}{Z}\prod_C \phi_C$ and let $i, j \in V$ not connected by an edge.

$$\text{Then } f(x_i, x_j, x_R) = \tfrac{1}{Z}(\textstyle\prod_{i \in C} \phi_C)(\prod_{j \in C} \phi_C)(\prod_{i,j \notin C} \phi_C)$$

Hence for all x_i, y_i, x_j, y_j, x_R: $f(x_i, x_j, x_R) f(y_i, y_j, x_R) = f(x_i, y_j, x_R) f(y_i, x_j, x_R)$

Average over the y_i, y_j: get $\frac{f(x_i, x_j, x_R)}{f(x_i, x_R)} = \frac{f(x_j, x_R)}{f(x_R)}$

$\Rightarrow$: Let f satisfy the pairwise Markov property w.r.t. G, Let $y \in \mathcal{X}$ arbitrary.

$$C \subset V \rightsquigarrow \phi_C(x_C) := \textstyle\prod_{S \subseteq C} f(x_S, y_{V \setminus S})^{(-1)^{|C|-|S|}}$$

Möbius inversion on the power set of V gives

$$f(x) = \prod_{C \subseteq V} \phi_C(x_C)^{\mu(V,C)}$$

It suffices to show: $\phi_C \equiv 1$ if C is not a clique

For this, choose $i, j \in C$ not connected by an edge, write down ϕ_C, and use the Markov property of i and j.

Corollary

Let P be a distribution that factors according to G. Then P satisfies the global Markov property on G.

Proof: the global Markov property is a closed condition and the statement is correct when P has positive density.

Parametric directed graphical model

All densities f with $f(x) = \prod_{j \in V} f(x_j | x_{\mathrm{pa}(j)})$

"Recursive Factorization Property"

Idea: we always have $f(x) = f(x_1)f(x_2|x_1)f(x_3|x_1, x_2) \cdots f(x_n|x_1, \ldots, x_{n-1})$

Here, the ordering of the vertices respects parenthood.

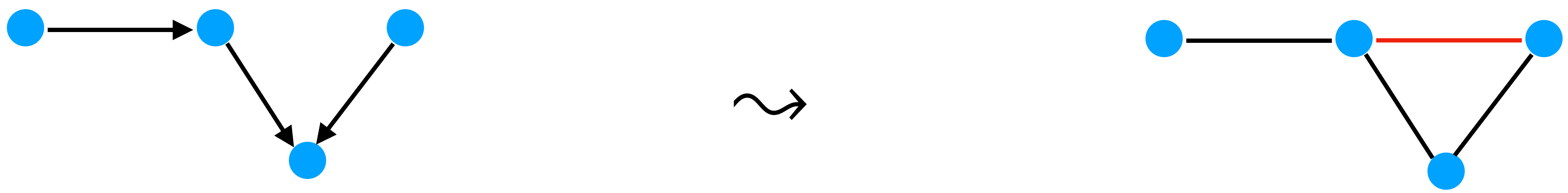
But the graph says that the information from the parents suffices.

Theorem 13.2.10 (Recursive Factorization). *A probability density satisfies the recursive factorization property (13.2.2) associated to the directed acyclic graph G if and only if it satisfies the directed local Markov property associated to G.*

Proof. (⇒) Let f factorize. Then it satisfies the *global* Markov property

Indeed, let C d-separate A, B, W.l.o.g. $V = \mathrm{an}(A \cup B \cup C)$

Then C separates A and B in the *moralization* G^{mor} of G



Moralization makes $\{j\} \cup \mathrm{pa}(j)$ into a clique, hence f factorizes according to G^{mor}

By the Corollary, $X_A \perp\!\!\!\perp X_B | X_C$

(⇐) carry out the Idea $f(x) = f(x_1) f(x_2|x_1) \cdots f(x_n|x_1, \ldots, x_n)$

Theorem

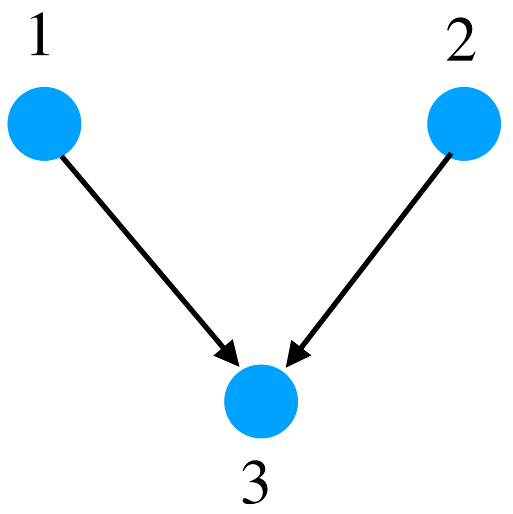
For any random variable X, directed graph G:

Local directed Markov w.r.t. $G \Rightarrow$ Global directed Markov w.r.t. G

$\Searrow$ $\Nearrow$

Recursive factorization property w.r.t. G

Example: discrete case



$X_i \in \{0, 1\}$

$$f(x_1, x_2, x_3) = f(x_1)f(x_2)f(x_3|x_1, x_2)$$

$$\begin{aligned}
p_{0,0,0} &= \theta_0^{(1)}\theta_0^{(2)}\theta_{0|0,0}^{(3)} & p_{1,0,0} &= \theta_1^{(1)}\theta_0^{(2)}\theta_{0|1,0}^{(3)} \\
p_{0,0,1} &= \theta_0^{(1)}\theta_0^{(2)}\theta_{1|0,0}^{(3)} & p_{1,0,1} &= \theta_1^{(1)}\theta_0^{(2)}\theta_{1|1,0}^{(3)} \\
p_{0,1,0} &= \theta_0^{(1)}\theta_1^{(2)}\theta_{0|0,1}^{(3)} & p_{1,1,0} &= \theta_1^{(1)}\theta_1^{(2)}\theta_{0|1,1}^{(3)} \\
p_{0,1,1} &= \theta_0^{(1)}\theta_1^{(2)}\theta_{1|0,1}^{(3)} & p_{1,1,1} &= \theta_1^{(1)}\theta_1^{(2)}\theta_{1|1,1}^{(3)}
\end{aligned}$$

$$\Delta_1 \times \Delta_1 \times \Delta_1^4 \to \Delta_7$$

Multivariate Gaussian case

X multivariate Gaussian $\Rightarrow$ X_i univariate Gaussian

$X_j | X_{\mathrm{pa}(j)}$ multivariate Gaussian

$$f(x) = \prod_j f(x_j | x_{\mathrm{pa}(j)}) \Rightarrow X_i = \sum_{j \in \mathrm{pa}(j)} \lambda_{i,j} X_j + \varepsilon_i$$

Where $\varepsilon_i \sim \mathcal{N}(\nu_i, \omega_i)$

We have $X = (\mathrm{Id} - \Lambda)^{-T} \varepsilon$

Where $\Lambda_{i,j} = \lambda_{i,j}$ if $(i,j) \in E$, 0 else.

$$\Rightarrow \Sigma = (\mathrm{Id} - \Lambda)^{-T} \Omega (\mathrm{Id} - \Lambda)^{-1}$$

With $\Omega = \mathrm{diag}(\omega_1, \ldots, \omega_n)$

Proposition 13.2.12. *The parametrized Gaussian graphical model associated to the directed acyclic graph G consists of all pairs $(\mu, \Sigma) \in \mathbb{R}^m \times PD_m$ such that $\Sigma = (Id - \Lambda)^{-T}\Omega(Id - \Lambda)^{-1}$ for some Ω diagonal with positive entries and upper triangular $\Lambda \in \mathbb{R}^E$.*

$$\mathcal{M}_{\text{paramGaussian}} = \text{image of } (\Lambda, \Omega) \mapsto \Sigma$$

Ideal of the closure: I_G

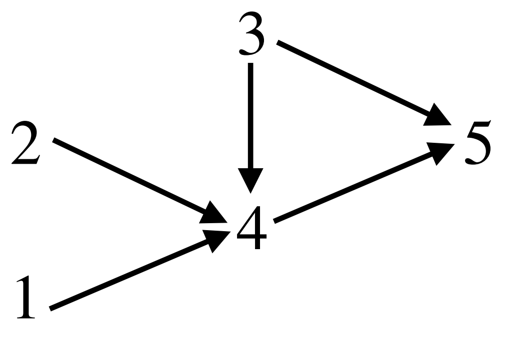
Ideal I_{glob} of conditional independent statements for $X \sim \mathcal{N}(\mu, \Sigma)$:

Ideal $I_{\text{glob}} = \sum_{A \perp\!\!\!\perp_d B|C} I_{A \perp\!\!\!\perp B|C}$

$I_{A \perp\!\!\!\perp B|C} = \langle (|C| + 1)\text{-minors of } \Sigma_{A \cup C, B \cup C} \rangle$

Question: when does $I_{\text{glob}} = I_G$?

Example

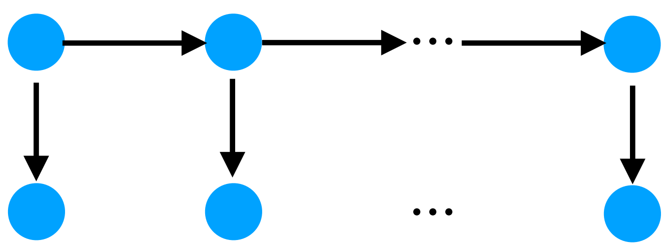


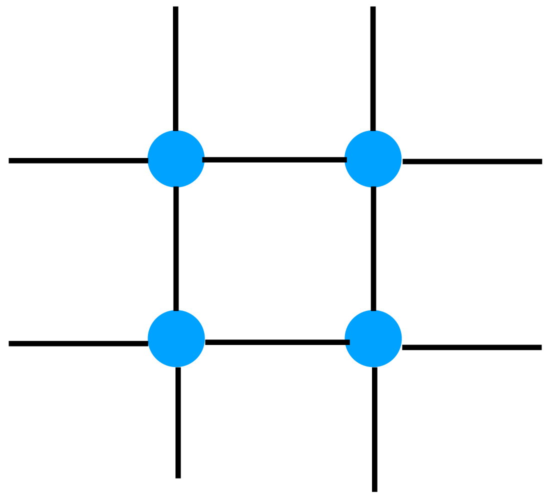
$$\det(\Sigma_{12,45}) \in I_G \setminus I_{\text{glob}}$$

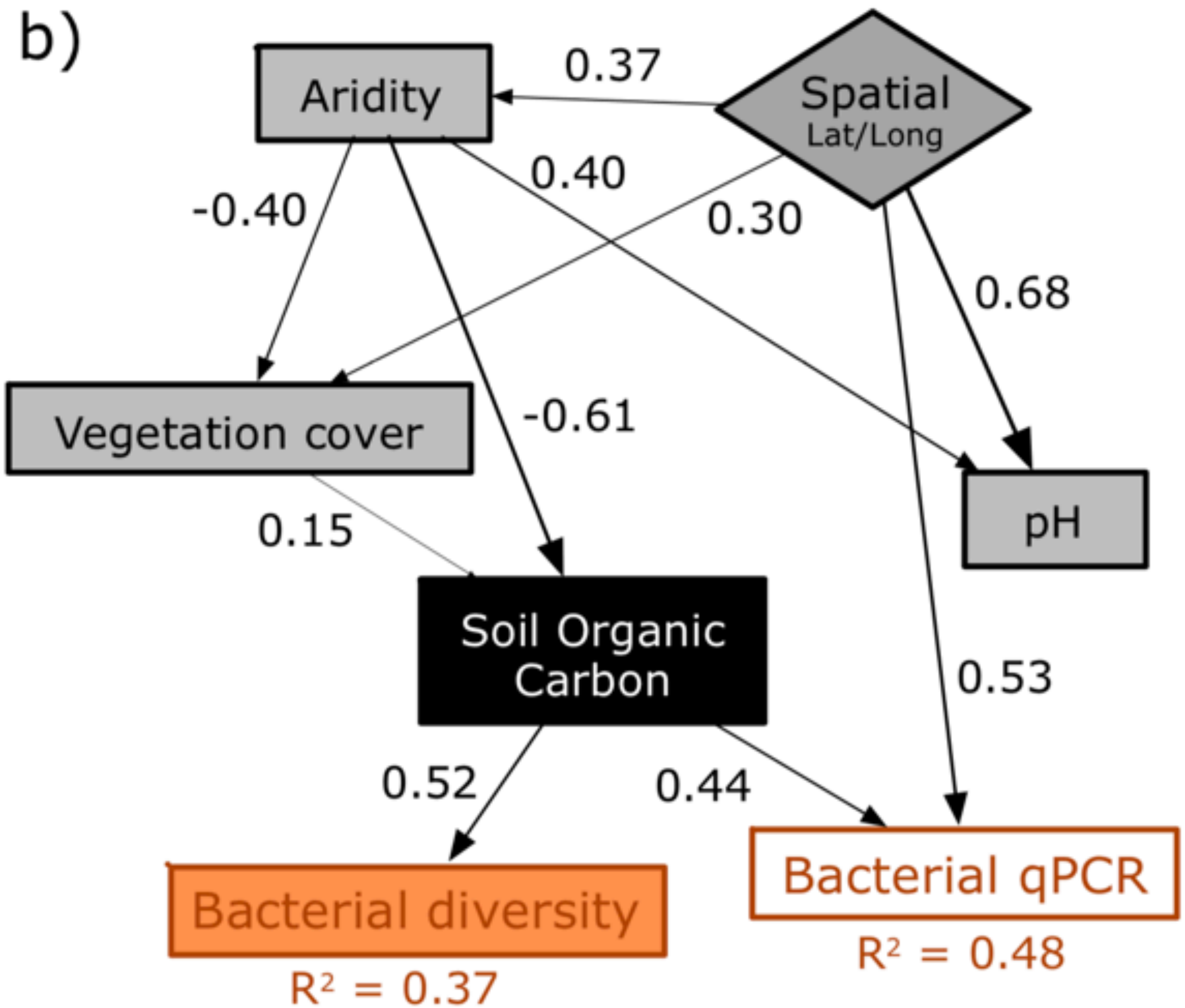
$$I_G = I_{\text{glob}} + \langle \det(\Sigma_{12,45}) \rangle$$

Examples of graphical models


Markov chain


Hidden Markov model


Ising model



(Guerra, Eisenhauer, Pereira: *Synthesising Soil Ecosystem Multifunctionality*)

Talk to Eliana or me about this!