# *Algebraic Classification of Small Bayesian Networks*

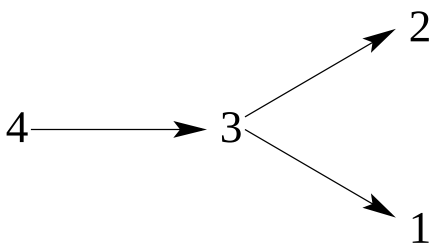
Luis David Garcia, Michael Stillman, and Bernd Sturmfels

`lgarcia@math.vt.edu`

Virginia Tech

# Bayesian Networks

- Let $G$ be a directed acyclic graph with $n$ nodes.
- The nodes represent random variables, denoted $X_1, \ldots, X_n$.
- The arrows represent causal dependencies among the variables.




- The joint probability distribution is defined as:

$$p(x_1, x_2, \ldots, x_n) = \prod_{i=1}^{n} p(x_i | x_{\mathrm{pa}_i})$$

- $p(x_1, x_2, x_3, x_4) = p(x_4)p(x_3|x_4)p(x_2|x_3)p(x_1|x_3)$

# Conditional Independence Relations

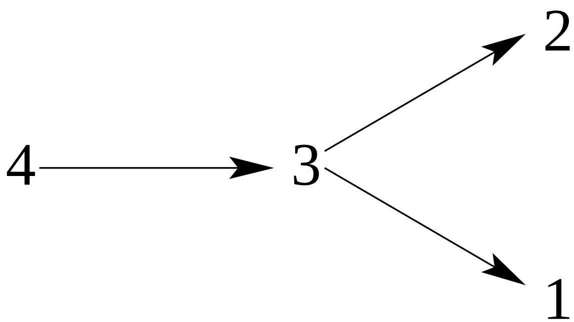
$$3 \longrightarrow 2 \longrightarrow 1 \qquad X_1 \perp\!\!\!\perp X_3 \mid X_2$$

$$2 \longleftarrow 3 \longrightarrow 1 \qquad X_1 \perp\!\!\!\perp X_2 \mid X_3$$

$$3 \longrightarrow 1 \longleftarrow 2 \qquad X_3 \perp\!\!\!\perp X_2$$

- The set of directed global Markov relations, $\mathrm{global}(G)$, is the set of independent statements $A \perp\!\!\!\perp B \mid C$, for any triple $A, B, C$ of disjoint subsets of vertices of $G$ such that $A$ and $B$ are d-separated by $C$.
- The set of directed local Markov relations of $G$ is the set of independence statements

$$\mathrm{local}(G) \quad = \quad \{X_i \perp\!\!\!\perp \mathrm{nd}(X_i) \mid \mathrm{pa}(X_i) : \; i = 1, 2, \ldots, n\},$$



Local: $\{X_1 \perp\!\!\!\perp \{X_2, X_4\} \mid X_3,\ X_2 \perp\!\!\!\perp \{X_1, X_4\} \mid X_3\}$

Global: *Local* $\cup \{X_4 \perp\!\!\!\perp \{X_1, X_2\} \mid X_3\}$

# *Ideals, Varieties, and Independent Statements*

- Let $X_1, \ldots, X_n$ be discrete random variables, where $X_i$ takes values in $[d_i] = \{1, 2, \ldots, d_i\}$.
- Let $D = [d_1] \times [d_2] \times \cdots \times [d_n]$ so that $\mathbb{R}^D$ denotes the real vector space of $n$-dimensional tables of format $d_1 \times \cdots \times d_n$.
- Let $p_{u_1 u_2 \cdots u_n}$ be an indeterminate representing the probability of $X_1 = u_1,\, X_2 = u_2, \ldots, X_n = u_n$.

# Ideals, Varieties, and Independent Statements (2)

- Let $A \perp\!\!\!\perp B \mid C$ be a conditional independence statement, where $A, B$ and $C$ are pairwise disjoint subsets of $\{X_1, \ldots, X_n\}$.
- Let $I_{A \perp\!\!\!\perp B | C}$ denote de ideal of $\mathbb{R}^D$ generated by the homogeneous quadratic polynomials

$$p(a,b,c)p(a',b',c) - p(a,b',c)p(a',b,c)$$

for every pair $a, a'$ of instances of $A$, every pair $b, b'$ of instances of $B$, and for every instance $c$ of $C$.

- $p(a,b,c)$ denotes de marginalization of $p(x_1, \ldots, x_n)$ over the variables in the complement of $A \cup B \cup C$.

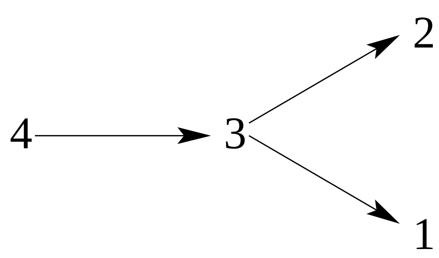
# Ideals, Varieties, and Independent Statements (3)

- Let $\mathcal{M}$ be the independence model
  $\mathcal{M} \quad = \quad \{A^{(1)} \perp\!\!\!\perp B^{(1)} \mid C^{(1)}, \ldots, A^{(m)} \perp\!\!\!\perp B^{(m)} \mid C^{(m)}\}$. Then

$$I_{\mathcal{M}} \quad = \quad I_{A^{(1)} \perp\!\!\!\perp B^{(1)} | C^{(1)}} + \cdots + I_{A^{(m)} \perp\!\!\!\perp B^{(m)} | C^{(m)}}$$

- The independence variety is the set $V(I_{\mathcal{M}})$ of common zeros in $\mathbb{C}^D$ of the polynomials in $I_{\mathcal{M}}$.
- $V(I_{\mathcal{M}})$ is the set of all $d_1 \times \cdots \times d_n$-tables with complex number entries which satisfy the conditional independence statements in $\mathcal{M}$.
- $V_{\geq}(I_{\mathcal{M}} + \langle p - 1 \rangle)$ is the subset of the probability simplex specified by the model $\mathcal{M}$. Where $p$ denotes the sum of all unknowns.

# *Ideals, Varieties, and Independent Statements (4)*




$$\text{local}(G) = \{X_1 \perp\!\!\!\perp \{X_2, X_4\} \mid X_3,\ X_2 \perp\!\!\!\perp \{X_1, X_4\} \mid X_3\}$$

$$\text{global}(G) = \text{local}(G) \cup \{X_4 \perp\!\!\!\perp \{X_1, X_2\} \mid X_3\}$$

- Let $d_1 = d_2 = d_4 = 2$ and $d_3$ arbitrary
- The ideal $I_{\text{local}(G)} = I_{\text{global}(G)}$ is generated by the $2 \times 2$-subdeterminants of the following $d_3$ matrices

$$\begin{pmatrix} p_{11k1} & p_{11k2} & p_{12k1} & p_{12k2} \\ p_{21k1} & p_{21k2} & p_{22k1} & p_{22k2} \end{pmatrix} \qquad \begin{pmatrix} p_{11k1} & p_{11k2} & p_{21k1} & p_{21k2} \\ p_{12k1} & p_{12k2} & p_{22k1} & p_{22k2} \end{pmatrix}$$

- For each $k \in [d_3]$, the corresponding quadratic binomials define the Segre Variety $\mathbb{P}^1 \times \mathbb{P}^1 \times \mathbb{P}^1 \longrightarrow \mathbb{P}^7$
- $V(I_{\mathcal{M}})$ is the join of $d_3$ Segre varieties $\mathbb{P}^1 \times \mathbb{P}^1 \times \mathbb{P}^1 \longrightarrow \mathbb{P}^7$.

**Theorem (Factorization Theorem).** *Let $G$ be a directed acyclic graph and $P$ a probability distribution on $V(G)$, the following are equivalent:*

*DF:* $P$ *admits a recursive factorization according to* $G$

*DG:* $P$ *obeys the Directed Global Markov Property*

*DL:* $P$ *obeys the Directed Local Markov Proterty*

- For any integer $r \in [n]$ and $u_i \in [d_i]$, denote the marginalization over the first $r$ random variables as:

$$p_{++\cdots+u_{r+1}\cdots u_n} \quad := \quad \sum_{i_1=1}^{d_1}\sum_{i_2=1}^{d_2}\cdots\sum_{i_r=1}^{d_r} p_{i_1 i_2\cdots i_r u_{r+1}\cdots u_n}.$$

- Denote by $\mathbf{p}$ the product of all of these linear forms.
- The equation of $\mathbf{p} = 0$ defines a hyperplane arrangement in $\mathbb{R}^D$.
- $I_{\mathrm{local}(G)}$ is prime locally outside this hyperplane arrangement, and hence so is $I_{\mathrm{global}(G)}$.

# Factorization Theorem (2)

**Theorem.** $\left(I_{\mathrm{local}(G)} : \mathbf{p}^\infty\right) \;=\; \left(I_{\mathrm{global}(G)} : \mathbf{p}^\infty\right) \;=\; \ker(\Phi).$

- The prime ideal $\ker(\Phi)$ is called the distinguished component.
- It is the set of all homogeneous polynomial functions on $\mathbb{R}^D$ which vanish on all probabilitity distributions that factor according to $G$.

**Theorem.** *The following four subsets of the probability simplex coincide:*

$$\begin{aligned} V_{\geq}(I_{\mathrm{local}(G)} + \langle p - 1\rangle) &= V_{\geq}(I_{\mathrm{global}(G)} + \langle p - 1\rangle) \\ = \quad V_{\geq}(\mathrm{kernel}(\Phi)) &= \mathrm{image}(\phi_{\geq}). \end{aligned}$$

# Bayesian Networks on three random variables

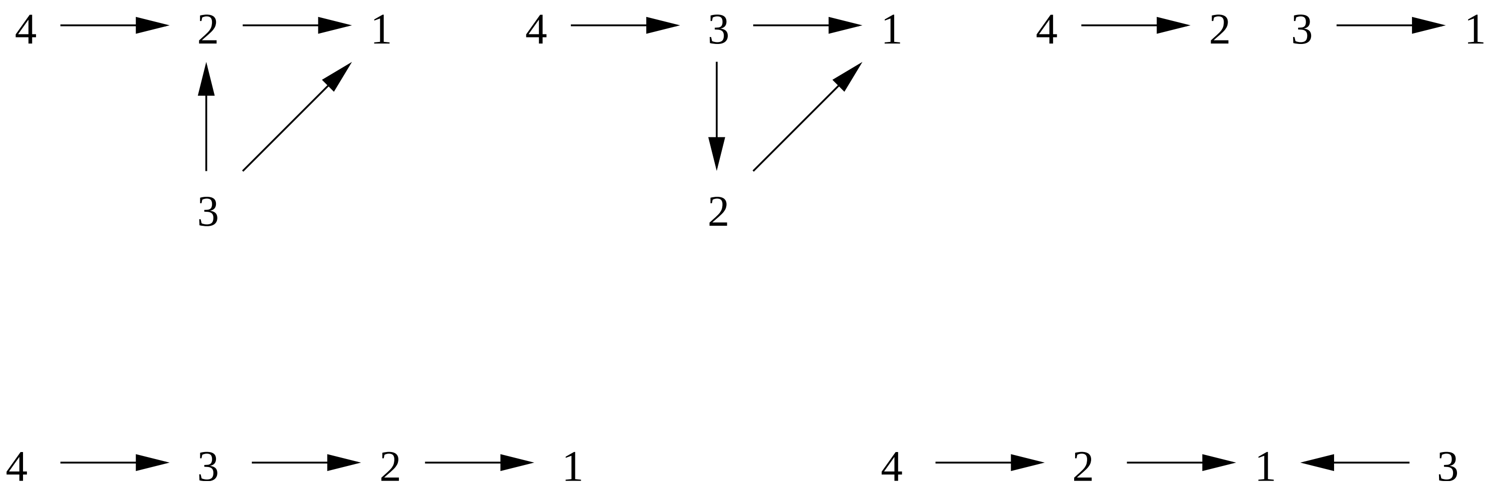
**Theorem.** *For any Bayesian network $G$ on three discrete random variables, the ideal $I_{\mathrm{local}(G)}$ is prime, and it has a quadratic Gröbner basis.*

| Graph | Local/Global Markov property | Independence ideal |
|---|---|---|
| $3 \quad 2 \quad 1$ | $1 \perp\!\!\!\perp \{2,3\},\ 2 \perp\!\!\!\perp \{1,3\},\ 3 \perp\!\!\!\perp \{1,2\}$ | $I_{\mathrm{Segre}}$ |
| $3 \longrightarrow 2 \quad 1$ | $1 \perp\!\!\!\perp 3,\ 1 \perp\!\!\!\perp 2 \mid 3,\ 1 \perp\!\!\!\perp \{2,3\}$ | $I_{1 \perp\!\!\!\perp \{2,3\}}$ |
| $3 \longrightarrow 2 \longrightarrow 1$ | $1 \perp\!\!\!\perp 3 \mid 2$ | $I_{1 \perp\!\!\!\perp 3 \mid 2}$ |
| $1 \longleftarrow 3 \longrightarrow 2$ | $1 \perp\!\!\!\perp 2 \mid 3$ | $I_{1 \perp\!\!\!\perp 2 \mid 3}$ |
| $3 \longrightarrow 1 \longleftarrow 2$ | $2 \perp\!\!\!\perp 3$ | $I_{2 \perp\!\!\!\perp 3}$ |

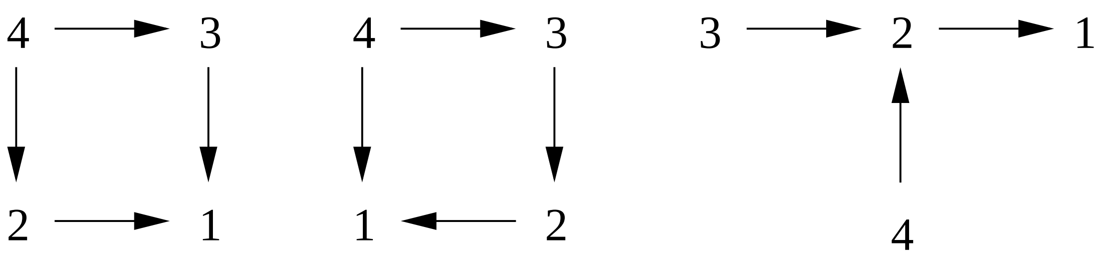
Virginia Tech

# Bayesian Networks on four random variables

**Theorem.** *Of the $30$ local Markov ideals on four random variables, $22$ are always* *prime, five are not prime but always radical*



*and three are not radical.*

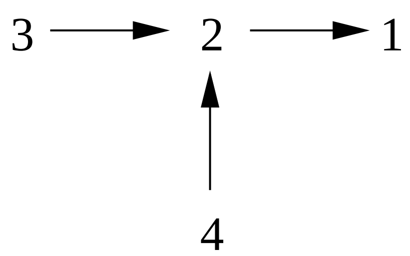


# ***A Prime example:*** $2 \longleftarrow 4 \longrightarrow 1 \longleftarrow 3$

- $\mathrm{Local}(G) = \{1 \perp\!\!\!\perp 2 \,|\, \{3,4\},\ 2 \perp\!\!\!\perp \{1,3\} \,|\, 4,\ 3 \perp \{2,4\}\}$.
- $I_{\mathrm{local}(G)}$ is binomial in $p_{ijkl}$ with $i \in \{+, 2, \ldots, d_1\}$.
- The generators are $p_{i_1 j_2 k_1 l} p_{i_2 j_1 k_2 l} - p_{i_1 j_1 k_1 l} p_{i_2 j_2 k_2 l}$, and $p_{+j_1 k_2 l_1} p_{+j_2 k_1 l_2} - p_{+j_1 k_1 l_1} p_{+j_2 k_2 l_2}$.
- The S-pairs within each group reduce to zero by the Gröbner basis property of the $2 \times 2$-minors of a generic matrix.
- The given set of irreducible quadrics is a reverse lexicographic Gröbner basis.
- The lowest variable is not a zero-divisor, and hence by symmetry none of the variables $p_{ijkl}$ is a zero-divisor.
- Since $\left(I_{\mathrm{local}(G)} : \mathbf{p}^{\infty}\right) = \ker(\Phi)$, then $I_{\mathrm{local}(G)}$ coincides with the prime ideal $\ker(\Phi)$.

- $I_{\mathrm{local}(G)}$ is the irredundant intersection of $2^{d_2} - 1$ primes.
- $\mathrm{local}(G) = \{1 \perp\!\!\!\perp \{3,4\} \,|\, 2,\ 2 \perp\!\!\!\perp 4 \,|\, 3\}$.
- $I_{\mathrm{local}(G)}$ is a binomial ideal in $p_{ijkl}$ with $i \in \{+, 2, 3, \ldots, d_1\}$.
- The minimal primes are indexed by proper subsets of $[d_2]$.
- For each subset $\sigma$ we introduce the monomial prime $M_\sigma = \langle p_{+jkl} \,:\, j \in \sigma, k \in [d_3], l \in [d_4] \rangle$
- The complementary monomial $m_\sigma = \prod_{j \in [d_2] \setminus \sigma} \prod_{k \in [d_3]} \prod_{l \in [d_4]} p_{+jkl}$,
- The ideal $P_\sigma = \big( (I_{\mathrm{local}(G)} + M_\sigma) \,:\, m_\sigma^\infty \big)$.
- These ideals are prime, and the union of their varieties is irredundant and equals the variety of $I_{\mathrm{local}(G)}$.

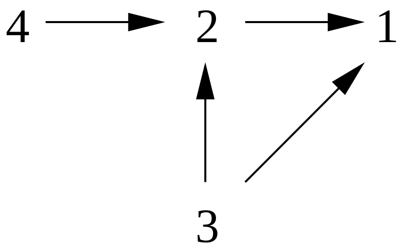
# A Non-radical example: Network 21



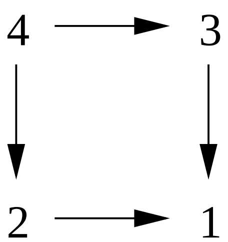

- $\mathrm{Local}(G) = \big\{1 \perp\!\!\!\perp \{3,4\} \,|\, 2,\ 3 \perp\!\!\!\perp 4\big\}$.
- $I_{\mathrm{local}(G)}$ is generated by $p_{i_1 j k_2 l_2} p_{i_2 j k_1 l_1} - p_{i_1 j k_1 l_1} p_{i_2 j k_2 l_2}$, and $p_{++k_1 l_2} p_{++k_2 l_1} - p_{++k_1 l_1} p_{++k_2 l_2}$.
- Let $d_1 = d_2 = d_3 = 2$ and $d_4 = 3$. $I_{\mathrm{local}(G)}$ is generated by $33$ quadratic polynomials in $24$ unknowns.
- The degree reverse lexicographic Gröbner basis of this ideal consists of $123$ polynomials of degree up to $8$.
- $I_{\mathrm{local}(G)}$ is the intersection of the distinguished component and the $P$-primary ideal $Q = I_{1 \perp\!\!\!\perp \{3,4\} \,|\, 2} + P^2$, where $P$ is the prime ideal generated by the $12$ linear forms $p_{+jkl}$.

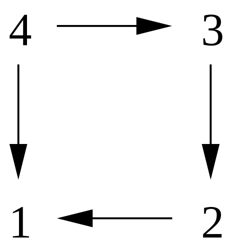
# Global Markov Relations on Four nodes

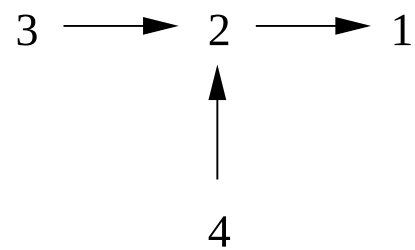
**Theorem.** *Of the $30$ global Markov ideals on four random variables, $26$ are always prime, one is not prime but always radical*




*and three are not radical*



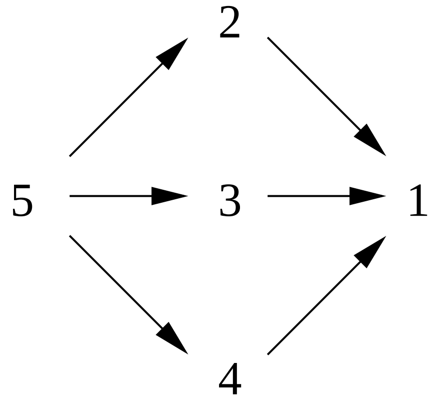



# Global Markov Relations on Five nodes

**Theorem.** *Of the* $301$ *global Markov ideals on five binary random variables,* $220$ *are prime,* $68$ *are radical but not prime, and* $13$ *are not radical.*

| # of components | 1 | 3 | 5 | 7 | 17 | 25 | 29 | 33 | 39 |
|---|---|---|---|---|---|---|---|---|---|
| # of ideals | 220 | 8 | 41 | 3 | 9 | 1 | 2 | 3 | 1 |

- `http://math.cornell.edu/~mike/bayes/global5.html`.




- $I_{\mathrm{global}(G_{138})}$ has $207$ minimal primes, and $37$ embedded primes. Each of the $207$ minimal primary components are prime.

# Hidden Variables and Higher Secant Varieties

- Let $G$ be a BN on $n$ discrete random variables and let $P_G = \ker(\Phi)$ be its homogeneous prime ideal.
- The variables corresponding to the nodes $r+1, \ldots, n$ are hidden,
- The *observable probabilities* are $p_{i_1 i_2 \cdots i_r ++\cdots+} = \sum_{j_{r+1} \in [d_{r+1}]} \sum_{j_{r+2} \in [d_{r+2}]} \cdots \sum_{j_n \in [d_n]} p_{i_1 i_2 \cdots i_r\, j_{r+1} j_{r+2} \cdots j_n}$.
- Let $D' = [d_1] \times \cdots \times [d_r]$ and $\mathbb{R}[D'] \subset \mathbb{R}[D]$ generated by $p_{i_1 i_2 \cdots i_r ++\cdots+}$.
- Let $\pi : \mathbb{R}^D \to \mathbb{R}^{D'}$ denote the canonical linear epimorphism induced by the inclusion of $\mathbb{R}[D']$ in $\mathbb{R}[D]$.
- $\pi(V_{\geq 0}(P_G)) \subset \pi(V(P_G))_{\geq 0} \subset \pi(V(P_G)) \subset \overline{\pi(V(P_G))} \subset \mathbb{R}^{D'}$.