Introduction to Algebraic Methods in Graphical Models

at the SIAM Conference on Applied Algebraic Geometry (AG23)

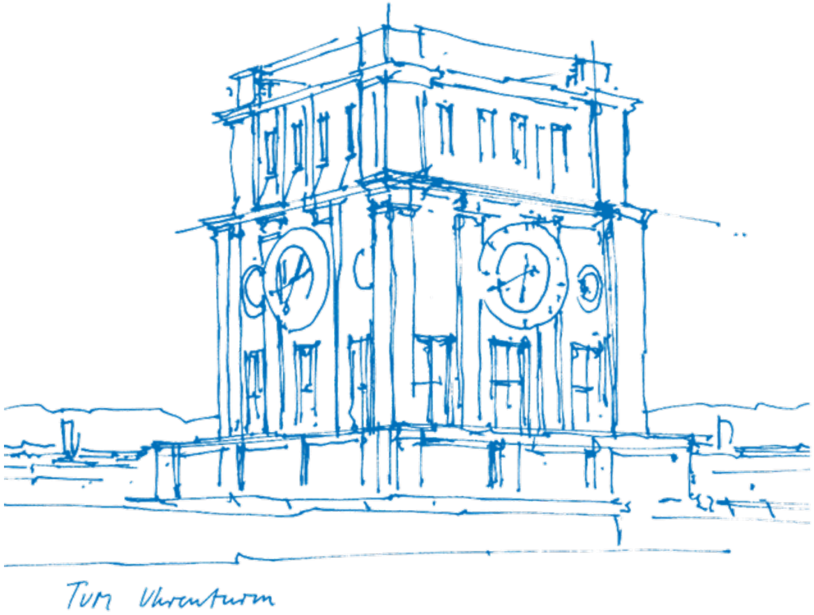
Nils Sturma

Research group Mathematical Statistics

TUM School of Computation, Information and Technology

Technical University of Munich

(joint work with Mathias Drton, Alexandros Grosdos and Irem Portakal)



Statistical Graphical Models

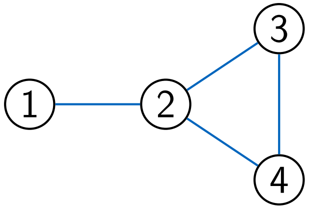
Setup: Random variables $(X_v)_{v\in V}$ and undirected/directed graph $G=(V,E)$.

$$\mathcal{M}(G) = \{\textit{probability distributions on } \mathbb{R}^{|V|} \textit{ that factorize according to } G\}$$

Undirected Graphical Models:

$$p(x) = \frac{1}{Z}\prod_{C\in\mathcal{C}} \psi_C(x_C),$$

where $\mathcal{C}$ is the collection of cliques of G.



$$p(x_1,x_2,x_3,x_4) \propto \psi_{12}(x_1,x_2)\cdot\psi_{234}(x_2,x_3,x_4)$$

Statistical Graphical Models

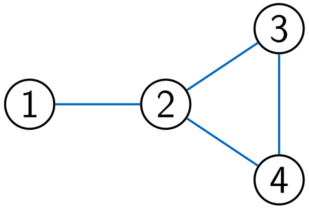
Setup: Random variables $(X_v)_{v \in V}$ and undirected/directed graph $G = (V, E)$.

$$\mathcal{M}(G) = \{\textit{probability distributions on } \mathbb{R}^{|V|} \textit{ that factorize according to } G\}$$

Undirected Graphical Models:

$$p(x) = \frac{1}{Z} \prod_{C \in \mathcal{C}} \psi_C(x_C),$$

where $\mathcal{C}$ is the collection of cliques of G.

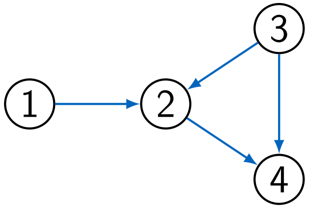


$$p(x_1, x_2, x_3, x_4) \propto \psi_{12}(x_1, x_2) \cdot \psi_{234}(x_2, x_3, x_4)$$

Directed Graphical Models:

$$p(x) = \prod_{v \in V} p(x_v | x_{\mathrm{pa}(v)}),$$

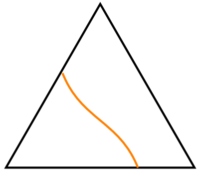
where $\mathrm{pa}(v) = \{w \in V : w \to v \in E\}$.



$$p(x_1, x_2, x_3, x_4) = p(x_1) \cdot p(x_2 | x_1, x_3) \cdot p(x_3) \cdot p(x_4 | x_2, x_3)$$

Polynomial Parameterizations in Directed Graphical Models

Discrete Distributions



- Finite state space:

$$\mathcal{I} = \times_{v \in V} [d_v].$$

- Each prob. distribution is a point in

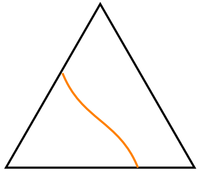
$$\Delta_{|\mathcal{I}|-1} = \{p \in \mathbb{R}^{\mathcal{I}} : p(x) \geq 0, \sum_{i \in \mathcal{I}} p(x) = 1\}.$$

- Graphical model $\mathcal{M}(G) \subseteq \Delta_{|\mathcal{I}|-1}$ is given by

$$\mathcal{M}(G) = \{p \in \mathbb{R}^{\mathcal{I}} : p(x) = \prod_{v \in V} p(x_v | x_{\mathrm{pa}(v)})\}.$$

Polynomial Parameterizations in Directed Graphical Models

Discrete Distributions



- Finite state space:

$$\mathcal{I} = \times_{v\in V}[d_v].$$

- Each prob. distribution is a point in

$$\Delta_{|\mathcal{I}|-1} = \{p \in \mathbb{R}^{\mathcal{I}} : p(x) \geq 0, \sum_{i\in\mathcal{I}} p(x) = 1\}.$$

- Graphical model $\mathcal{M}(G) \subseteq \Delta_{|\mathcal{I}|-1}$ is given by

$$\mathcal{M}(G) = \{p \in \mathbb{R}^{\mathcal{I}} : p(x) = \prod_{v\in V} p(x_v|x_{\mathrm{pa}(v)})\}.$$

Gaussian Distributions

- Each distribution $N(0,\Sigma)$ is given by cov. matrix

$$\Sigma \in PD(|V|) \subseteq \mathbb{R}^{\binom{|V|+1}{2}}.$$

- Weight matrix:

$$\mathbb{R}^E = \{\Lambda \in \mathbb{R}^{|V|\times|V|} : \lambda_{vu} = 0 \text{ if } u \to v \notin E\}.$$

- Graphical model $\mathcal{M}^{(2)}(G) \subseteq PD(|V|)$ is given by

$$\mathcal{M}^{(2)}(G) = \{\Sigma \in PD(|V|) : \Sigma = (I-\Lambda)^{-1}\Omega(I-\Lambda)^{-\top}\},$$

where $\Lambda \in \mathbb{R}^E, \Omega \in \mathrm{diag}_+$.

Multiple graphical models have a polynomial parameterization.

Example: Gaussian Directed Graphical Models

$$\Lambda = \begin{pmatrix} 0 & 0 & 0 \\ \lambda_{21} & 0 & 0 \\ 0 & \lambda_{32} & 0 \end{pmatrix} \qquad \Omega = \begin{pmatrix} \omega_{11} & 0 & 0 \\ 0 & \omega_{22} & 0 \\ 0 & 0 & \omega_{33} \end{pmatrix}$$

Covariance matrix:

$$\Sigma = \begin{pmatrix} \sigma_{11} & \sigma_{12} & \sigma_{13} \\ \cdot & \sigma_{22} & \sigma_{23} \\ \cdot & \cdot & \sigma_{33} \end{pmatrix} = \begin{pmatrix} \omega_{11} & \omega_{11}\lambda_{21} & \omega_{11}\lambda_{21}\lambda_{32} \\ \cdot & \omega_{22} + \omega_{11}\lambda_{21}^2 & \omega_{22}\lambda_{32} + \omega_{11}\lambda_{21}^2\lambda_{32} \\ \cdot & \cdot & \omega_{33} + \omega_{22}\lambda_{32}^2 + \omega_{11}\lambda_{21}^2\lambda_{32}^2 \end{pmatrix}.$$

Graphical model:

$$\mathcal{M}^{(2)}(G) = PD(|V|) \cap \mathbb{V}(\sigma_{12}\sigma_{23} - \sigma_{22}\sigma_{13})$$

Questions of Interest in Graphical Modeling

Statistics

- Model selection.
- Parameter identifiability.
- Parameter estimation.
  (Maximum-Likelihood-Estimation.)
- ...

Algebraic Geometry

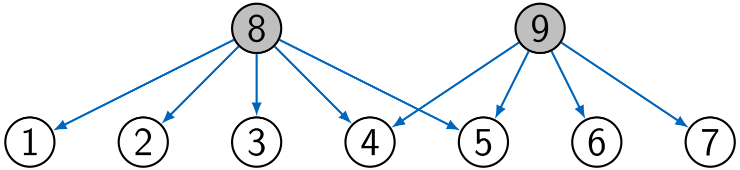
- Vanishing ideal of $\mathcal{M}(G)$.
- Is the parameterization map injective?
- Boundaries and singular locus of $\mathcal{M}(G)$.
  ML degree.
- ...

Interesting:

Hidden variables, i.e., $V = \mathcal{O} \cup \mathcal{H}$. Only observe marginal distribution $X_{\mathcal{O}}$.

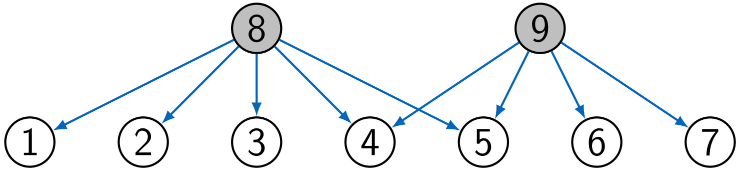
Factor Analysis Models

Graph $G = (\mathcal{O} \cup \mathcal{H}, E)$, only edges from latent to observed variables. Usually, $\mathcal{O} = [p]$.



Factor Analysis Models

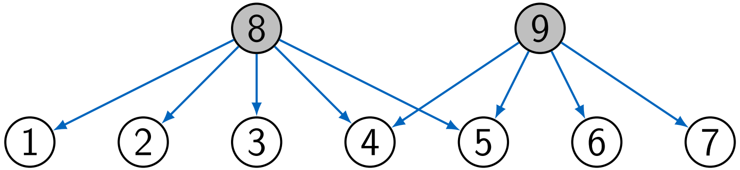
Graph $G = (\mathcal{O} \cup \mathcal{H}, E)$, only edges from latent to observed variables. Usually, $\mathcal{O} = [p]$.



$$\Lambda = \begin{pmatrix} 0 & \Lambda_{\mathcal{O}\mathcal{H}} \\ 0 & 0 \end{pmatrix} \qquad \Omega = \begin{pmatrix} \Omega_{\mathcal{O}\mathcal{O}} & 0 \\ 0 & \Omega_{\mathcal{H}\mathcal{H}} \end{pmatrix}$$

Factor Analysis Models

Graph $G = (\mathcal{O} \cup \mathcal{H}, E)$, only edges from latent to observed variables. Usually, $\mathcal{O} = [p]$.



$$\Lambda = \begin{pmatrix} 0 & \Lambda_{\mathcal{O}\mathcal{H}} \\ 0 & 0 \end{pmatrix} \qquad \Omega = \begin{pmatrix} \Omega_{\mathcal{O}\mathcal{O}} & 0 \\ 0 & \Omega_{\mathcal{H}\mathcal{H}} \end{pmatrix}$$

- Covariance matrix:

$$\mathrm{Cov}\left[\begin{pmatrix} X_{\mathcal{O}} \\ X_{\mathcal{H}} \end{pmatrix}\right] = (I - \Lambda)^{-1}\Omega(I - \Lambda)^{-\top} = \begin{pmatrix} \Omega_{\mathcal{O}\mathcal{O}} + \Lambda_{\mathcal{O}\mathcal{H}}\Omega_{\mathcal{H}\mathcal{H}}\Lambda_{\mathcal{O}\mathcal{H}}^\top & \Lambda_{\mathcal{O}\mathcal{H}}\Omega_{\mathcal{H}\mathcal{H}} \\ \Omega_{\mathcal{H}\mathcal{H}}\Lambda_{\mathcal{O}\mathcal{H}}^\top & \Omega_{\mathcal{H}\mathcal{H}} \end{pmatrix}$$

- *Observed* covariance matrix (projection):

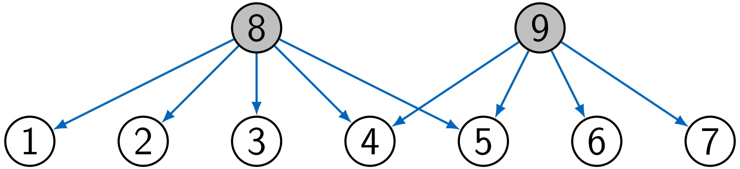
$$\mathrm{Cov}[X_{\mathcal{O}}] = \Omega_{\mathcal{O}\mathcal{O}} + \Lambda_{\mathcal{O}\mathcal{H}}\Omega_{\mathcal{H}\mathcal{H}}\Lambda_{\mathcal{O}\mathcal{H}}^\top$$

- *Observed* covariance model:

$$F_G = \{\widetilde{\Omega} + \widetilde{\Lambda}\widetilde{\Lambda}^\top \in \mathbb{R}^{|\mathcal{O}|\times|\mathcal{O}|} : \widetilde{\Omega} > 0 \text{ diagonal}, \ \widetilde{\Lambda} \in \mathbb{R}^{E_{\mathcal{O}\mathcal{H}}}\}$$

Goals: $\dim(F_G)$? $\quad I(F_G) \subseteq \mathbb{R}[\sigma_{ij}, i < j]$?

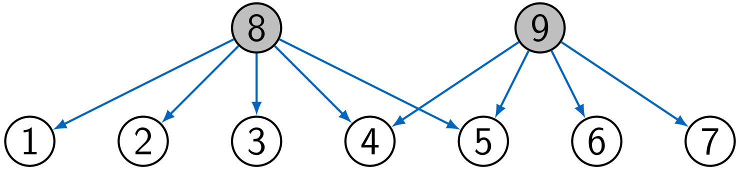
Example: Factor Analysis Models



Parameter matrix:

$$\tilde{\Lambda} = \begin{pmatrix} \lambda_{18} & \lambda_{28} & \lambda_{38} & \lambda_{48} & \lambda_{58} & 0 & 0 \\ 0 & 0 & 0 & \lambda_{49} & \lambda_{59} & \lambda_{69} & \lambda_{79} \end{pmatrix}^{\top}, \qquad \widetilde{\Omega} = \mathrm{diag}(\omega_{11}, \omega_{22}, \omega_{33}, \omega_{44}, \omega_{55}, \omega_{66}, \omega_{77}).$$

Example: Factor Analysis Models



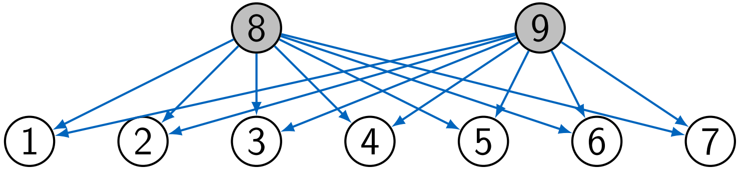
Parameter matrix:

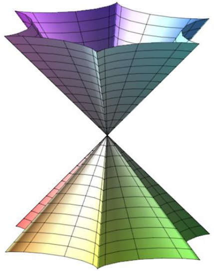
$$\widetilde{\Lambda} = \begin{pmatrix} \lambda_{18} & \lambda_{28} & \lambda_{38} & \lambda_{48} & \lambda_{58} & 0 & 0 \\ 0 & 0 & 0 & \lambda_{49} & \lambda_{59} & \lambda_{69} & \lambda_{79} \end{pmatrix}^{\top}, \qquad \widetilde{\Omega} = \mathrm{diag}(\omega_{11}, \omega_{22}, \omega_{33}, \omega_{44}, \omega_{55}, \omega_{66}, \omega_{77}).$$

Observed covariance matrix:

$$\Sigma = \begin{pmatrix} \omega_{11} + \lambda_{18}^2 & \lambda_{18}\lambda_{28} & \lambda_{18}\lambda_{38} & \lambda_{18}\lambda_{48} & \lambda_{18}\lambda_{58} & 0 & 0 \\ \lambda_{18}\lambda_{28} & \omega_{22} + \lambda_{28}^2 & \lambda_{28}\lambda_{38} & \lambda_{28}\lambda_{48} & \lambda_{28}\lambda_{58} & 0 & 0 \\ \lambda_{18}\lambda_{38} & \lambda_{28}\lambda_{38} & \omega_{33} + \lambda_{38}^2 & \lambda_{38}\lambda_{48} & \lambda_{38}\lambda_{58} & 0 & 0 \\ \lambda_{18}\lambda_{48} & \lambda_{28}\lambda_{48} & \lambda_{38}\lambda_{48} & \omega_{44} + \lambda_{48}^2 + \lambda_{49}^2 & \lambda_{48}\lambda_{58} + \lambda_{49}\lambda_{59} & \lambda_{49}\lambda_{69} & \lambda_{49}\lambda_{79} \\ \lambda_{18}\lambda_{58} & \lambda_{28}\lambda_{58} & \lambda_{38}\lambda_{58} & \lambda_{48}\lambda_{58} + \lambda_{49}\lambda_{59} & \omega_{55} + \lambda_{58}^2 + \lambda_{59}^2 & \lambda_{59}\lambda_{69} & \lambda_{59}\lambda_{79} \\ 0 & 0 & 0 & \lambda_{49}\lambda_{69} & \lambda_{59}\lambda_{69} & \omega_{66} + \lambda_{69}^2 & \lambda_{69}\lambda_{79} \\ 0 & 0 & 0 & \lambda_{49}\lambda_{79} & \lambda_{59}\lambda_{79} & \lambda_{69}\lambda_{79} & \omega_{77} + \lambda_{79}^2 \end{pmatrix}.$$

Previous work: *Full* Factor Analysis Models



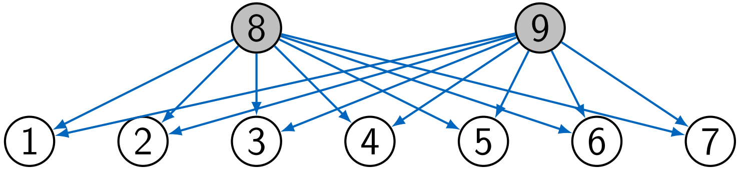


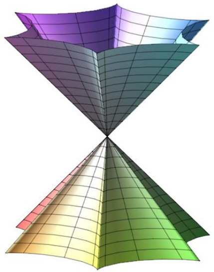
Dimension: [Drton et al., 2007]

$$\dim(F_G) = \min\{p(|\mathcal{H}|+1) - \binom{|\mathcal{H}|}{2}, \binom{p+1}{2}\},$$

where $|\mathcal{O}| = p$.

Previous work: *Full* Factor Analysis Models





Dimension: [Drton et al., 2007]

$$\dim(F_G) = \min\{p(|\mathcal{H}| + 1) - \binom{|\mathcal{H}|}{2}, \binom{p+1}{2}\},$$

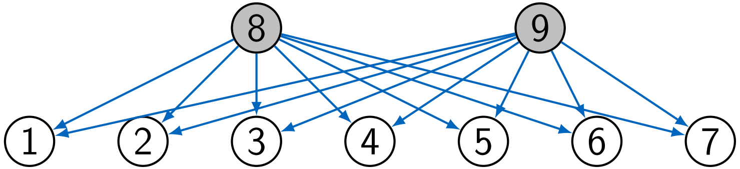
where $|\mathcal{O}| = p$.

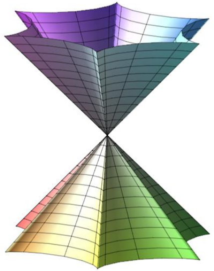
Ideal: [Drton et al., 2007]

$$I(F_G) = M_{p,|\mathcal{H}|} \cap \mathbb{R}[\sigma_{ij}, i < j],$$

where $M_{p,|\mathcal{H}|} \subseteq \mathbb{R}[\sigma_{ij}, i \leq j]$ is the ideal generated by all $(|\mathcal{H}| + 1) \times (|\mathcal{H}| + 1)$ minors of a symmetric $p \times p$ matrix.

Previous work: *Full* Factor Analysis Models





Dimension: [Drton et al., 2007]

$$\dim(F_G) = \min\{p(|\mathcal{H}|+1) - \binom{|\mathcal{H}|}{2}, \binom{p+1}{2}\},$$

where $|\mathcal{O}| = p$.

Ideal: [Drton et al., 2007]

$$I(F_G) = M_{p,|\mathcal{H}|} \cap \mathbb{R}[\sigma_{ij}, i < j],$$

where $M_{p,|\mathcal{H}|} \subseteq \mathbb{R}[\sigma_{ij}, i \leq j]$ is the ideal generated by all $(|\mathcal{H}|+1) \times (|\mathcal{H}|+1)$ minors of a symmetric $p \times p$ matrix.

Gröbner basis:

$|\mathcal{H}| = 1$:

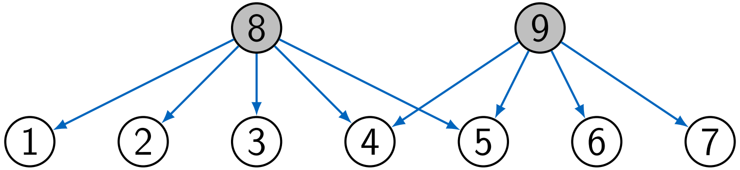
$$G_{p,1} = \{\sigma_{ij}\sigma_{kl} - \sigma_{ik}\sigma_{jl}, \sigma_{il}\sigma_{jk} - \sigma_{ik}\sigma_{jl}$$
$$1 \leq i < j < k < l \leq p\}.$$

$|\mathcal{H}| = 2$:

- $\overline{F_G}$ secant variety of the 1-factor model.
- Delightful strategy. [Sullivant, 2009]

New Project: *Sparse* Factor Analysis Models

At least one edge is missing, only $|\mathcal{H}| = 2$ latent variables.



Dimension:

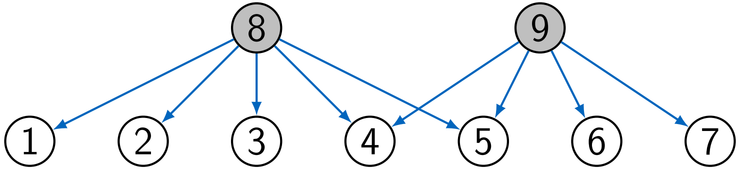
If $p \geq 5$ and any latent node has at least 3 children, then $\dim(F_G) = p + |E|$.

What about $I(F_G)$?

- If $\text{pa}(u) \cap \text{pa}(v) = \emptyset$, then $\sigma_{uv} = 0$ (singletons).
- Let $A, B \subseteq \mathcal{O}$ be disjoint sets s.t. $|A| = |B| = 2$. If $|\text{pa}(A) \cap \text{pa}(B)| \leq 1$, then $\det(\Sigma_{A,B}) = 0$ (tetrads).

New Project: *Sparse* Factor Analysis Models

At least one edge is missing, only $|\mathcal{H}| = 2$ latent variables.



Dimension:

If $p \geq 5$ and any latent node has at least 3 children, then $\dim(F_G) = p + |E|$.

What about $I(F_G)$?

- If $\mathrm{pa}(u) \cap \mathrm{pa}(v) = \emptyset$, then $\sigma_{uv} = 0$ (singletons).
- Let $A, B \subseteq \mathcal{O}$ be disjoint sets s.t. $|A| = |B| = 2$. If $|\mathrm{pa}(A) \cap \mathrm{pa}(B)| \leq 1$, then $\det(\Sigma_{A,B}) = 0$ (tetrads).

Theorem:

$$\mathbb{V}(I(F_G)) = \mathbb{V}(\{M_{p,2} + S^{\leq 1}(G)\} \cap \mathbb{R}[\sigma_{ij}, i < j]),$$

where $S^{\leq 1}(G)$ is the ideal generated by all singletons and tetrads corresponding to G.

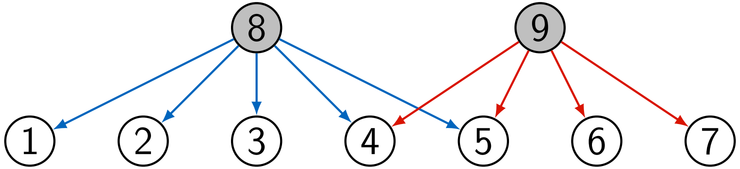
Joins

Join of Varieties:

$$W_1 * W_2 = \overline{\{\lambda w_1 + (1-\lambda) w_2 : w_1 \in W_1, w_2 \in W_2, \lambda \in \mathbb{R}\}}$$

(Sparse) Factor Models = Joins of (Sparse) One-Factor Models:

$$\sigma_{uv} = \begin{cases} \omega_{vv} + \sum_{h \in \mathrm{pa}(v)} \lambda_{vh}^2 & \text{if } u = v, \\ \sum_{h \in \mathrm{pa}(u) \cap \mathrm{pa}(v)} \lambda_{uh}\lambda_{vh} & \text{if } u \neq v. \end{cases}$$



<u>In accordance:</u> For two ideals I_1, I_2, one defines the join ideal $I_1 * I_2$ such that $I_1 * I_2 = I(\mathbb{V}(I_1) * \mathbb{V}(I_2))$.

$$\Longrightarrow \quad I(F_G) = I_1 * I_2$$

Delightful Strategy for Gröbner Basis

Observation:

$\text{in}_\prec(I_1 * I_2) \subseteq \text{in}_\prec(I_1) * \text{in}_\prec(I_2)$ for any term order $\prec$. If equality holds, then $\prec$ is *delightful* for I_1, I_2.

"Delightful" strategy: [Sturmfels, Sullivant, 2006]

Find $G \subseteq I_1 * I_2$ such that $\langle \text{in}_\prec(g) | g \in G \rangle = \text{in}_\prec(I_1) * \text{in}_\prec(I_2)$. Then G is a Gröbner basis.

1. Find delightful term order $\prec$.
2. Understand $\text{in}_\prec(I_1) * \text{in}_\prec(I_2)$.
3. Define polynomials $g \in G$ with correct initial terms.

Circular Term Order

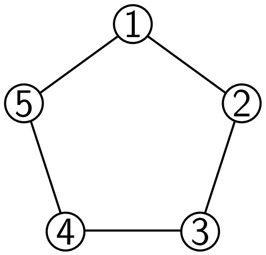
Circular distance on $\mathcal{O} = [p]$:

Length of the shortest path on regular p-gon.

Circular order:

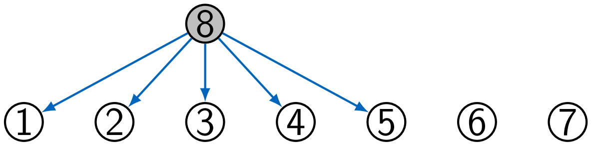
Any block-term order $\prec$ such that $\sigma_{uv} \succ \sigma_{wz}$ whenever the circular distance between u and v is smaller than the circular distance of w and z.

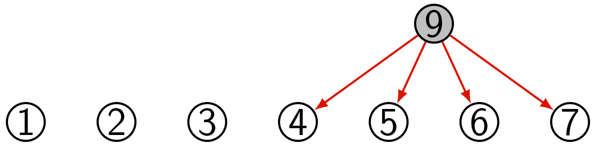
Example:



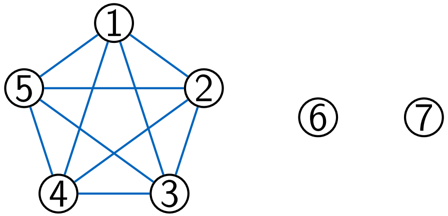
- $\sigma_{15} \succ \sigma_{24}$,
- $\sigma_{34} \succ \sigma_{25}$.

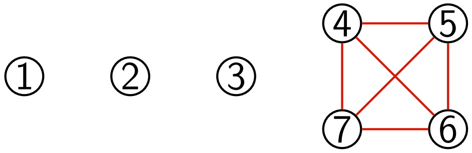
Non-Crossing Edge Graphs



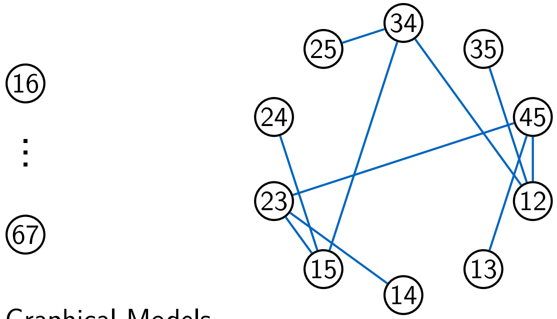


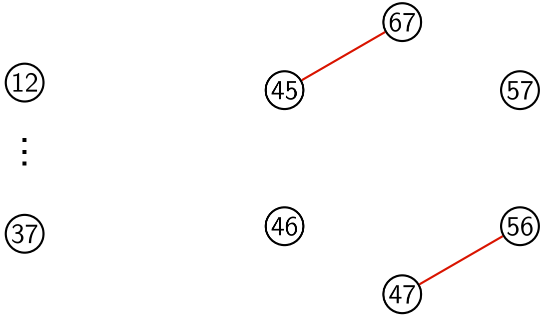
Complete graph on $v \in V$ s.t. $\mathrm{pa}(v) \neq \emptyset$:





Non-crossing edge graph:

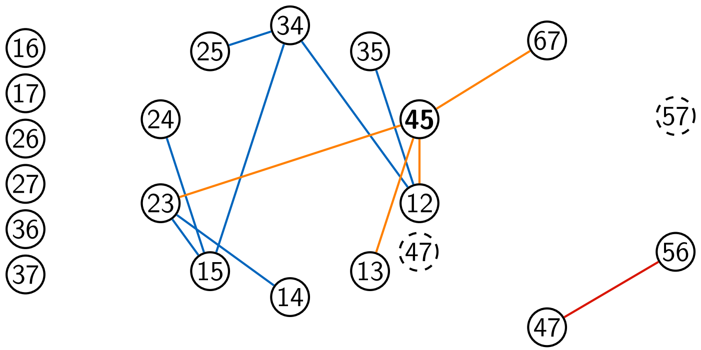




Gröbner Basis for "Overlap = 2"

$in_{\prec}(I_1) * in_{\prec}(I_2)$ is given by the edge ideal of the hypergraph obtained by gluing the non-crossing edge graphs.

Glued Hypergraph



Gröbner basis of $I_1 * I_2 = I(F_G)$

- $\sigma_{16}, \sigma_{17}, \sigma_{26}, \sigma_{27}, \sigma_{36}, \sigma_{37}$.
- $\sigma_{47}\sigma_{56} - \sigma_{57}\sigma_{46}$,
  $\sigma_{12}\sigma_{34} - \sigma_{13}\sigma_{24}$,
  $\sigma_{14}\sigma_{23} - \sigma_{13}\sigma_{24}$,
  $\sigma_{12}\sigma_{35} - \sigma_{13}\sigma_{25}$,
  $\sigma_{15}\sigma_{23} - \sigma_{13}\sigma_{25}$,
  $\sigma_{15}\sigma_{24} - \sigma_{14}\sigma_{25}$,
  $\sigma_{15}\sigma_{34} - \sigma_{14}\sigma_{35}$,
  $\sigma_{25}\sigma_{34} - \sigma_{24}\sigma_{35}$.
- $\sigma_{67}\sigma_{12}\sigma_{45} - \sigma_{67}\sigma_{24}\sigma_{15} - \sigma_{12}\sigma_{47}\sigma_{56}$,
  $\sigma_{67}\sigma_{13}\sigma_{45} - \sigma_{67}\sigma_{34}\sigma_{15} - \sigma_{13}\sigma_{47}\sigma_{56}$,
  $\sigma_{67}\sigma_{23}\sigma_{45} - \sigma_{67}\sigma_{34}\sigma_{25} - \sigma_{23}\sigma_{47}\sigma_{56}$.

Conclusion

- Latent variable models generally feature complicated geometry.
- Even in "simple" factor analysis models there is still lots to explore ...
- Circular term order is not delightful for "overlap $= 3$" ...
- What about $|\mathcal{H}| \geq 3$?

European Research Council
Established by the European Commission

Registration and travel support for this presentation was provided by the Society for Industrial and Applied Mathematics.

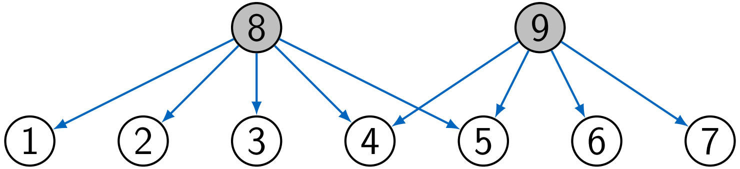
References

Drton, Sturmfels, Sullivant (2007).
Algebraic factor analysis: tetrads, pentads and beyond. Probab. Theory Relat. Fields 138:463-493.

Sturmfels, Sullivant (2006).
Combinatorial secant varieties. Pure Appl. Math. Q. 2(3):867-891.

Sullivant (2009).
A Gröbner basis for the secant ideal of the second hypersimplex. J. Commut. Algebra 1(2):327 – 338.

Appendix: Example



Singletons:

$\sigma_{16}, \quad \sigma_{26}, \quad \sigma_{36}, \quad \sigma_{17}, \quad \sigma_{27}, \quad \sigma_{37}.$

Tetrads:

- $A = \{1, 2\}$ and $B = \{4, 5\}$. Then $\text{pa}(A) \cap \text{pa}(B) = \{8\}$ and $\sigma_{14}\sigma_{25} - \sigma_{24}\sigma_{15} \in S^{\leq 1}(G)$.
- $A = \{1, 4\}$ and $B = \{2, 5\}$. Then $\text{pa}(A) \cap \text{pa}(B) = \{8, 9\}$ and $\sigma_{12}\sigma_{45} - \sigma_{24}\sigma_{15} \notin S^{\leq 1}(G)$.