Maximal Semigraphoids

Laura Silverstein and Dr. Michael O'Sullivan
Department of Mathematics and Statistics, San Diego State University

Purpose

Semigraphoids are combinatorial structures arising from models of probabilistic conditional independence. We investigate the classification of maximal semigraphoids on n random variables to better understand these sets.

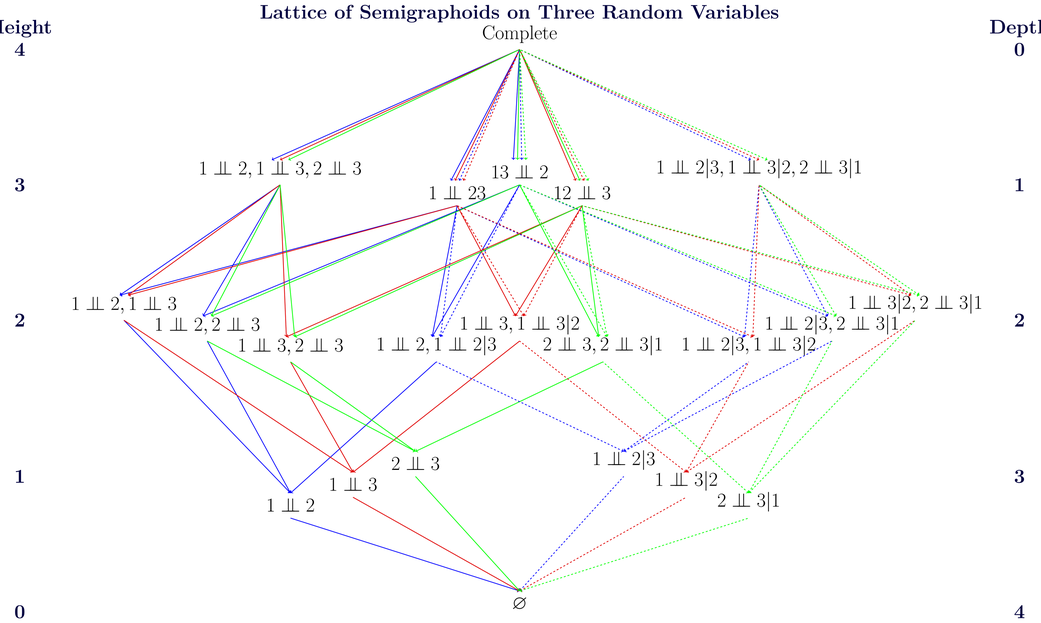
- Let A, B, C be pairwise disjoint subsets of the set $[n] = \{1, \ldots, n\}$. We regard statements of the form $A \perp\!\!\!\perp B|C$ as triplets over $[n]$.

What is a Semigraphoid?

A *semigraphoid* $\mathcal{S}$ is a set of triplets that satisfy the following for every pairwise disjoint $A, B, C, D \subset [n]$:

1. Symmetry: If $A \perp\!\!\!\perp B|C \in \mathcal{S}$, then $B \perp\!\!\!\perp A|C \in \mathcal{S}$.
2. Decomposition: If $A \perp\!\!\!\perp BC|D \in \mathcal{S}$, then $A \perp\!\!\!\perp B|D \in \mathcal{S}$.
3. Weak Union: If $A \perp\!\!\!\perp BC|D \in \mathcal{S}$, then $A \perp\!\!\!\perp B|CD \in \mathcal{S}$.
4. Contraction: If $\{A \perp\!\!\!\perp B|CD,\ A \perp\!\!\!\perp C|D\} \subset \mathcal{S}$, then $A \perp\!\!\!\perp BC|D \in \mathcal{S}$. [1,4]

Lattice of Semigraphoids on Three Random Variables



Semigraphoids as a Lattice

- With the binary relation $\subseteq$, the set of all semigraphoids on n random variables is a lattice [4]. Let $\mathcal{S}$ and $\mathcal{T}$ be semigraphoids. Then the join of $\mathcal{S}$ and $\mathcal{T}$ is $\langle \mathcal{S} \cup \mathcal{T} \rangle$ and the meet is precisely $\mathcal{S} \cap \mathcal{T}$.
- Statements of the form $a \perp\!\!\!\perp b|C$ where $a, b \in [n]$ and $C \subseteq [n] \setminus \{a, b\}$ are the building blocks of the lattice of semigraphoids on n random variables. They are the *atoms*.
- The lattice of semigraphoids is *atomic*. That is, every semigraphoid is generated by a join of atoms [3].
- We define an atom $a \perp\!\!\!\perp b|C$ to be a *k-atom* if $|C| = k$. The set $\{k - \text{atoms}\}$ is the set of all k-atoms in $CI(n)$ for a certain cardinality k.

Maximality of non-k-atoms

Theorem. If $A \subseteq \langle \{l - \text{atoms} : l < k\} \rangle$ and $B \subseteq \langle \{l - \text{atoms} : l > k\} \rangle$, then $\langle A \cup B \rangle = \langle A \rangle \cup \langle B \rangle$.

Corollary. Given any integer $k \leq n - 2$ and generating set of atoms $\mathcal{C}$. A semigraphoid $\mathcal{S} = \langle \mathcal{C} \rangle$ contains a k-atom if and only if there is at least one k-atom in $\mathcal{C}$.

Theorem. The semigraphoid $\langle \cup_{l \in \{1,\ldots,n-2\} \setminus k} \{l - \text{atoms}\} \rangle$ is maximal for any $k \in \{1, \ldots, n-2\}$. This semigraphoid is notated $\langle \text{non} - k - \text{atoms} \rangle$ and called the *non-k-atoms*.

- Let $a \perp\!\!\!\perp b|C$ be any k-atom for $k < n - 2$. Then we have $\langle a \perp\!\!\!\perp b|C \cup \{(k+1) - \text{atoms}\} \rangle = \langle \{x \perp\!\!\!\perp y|C : \text{for all } x, y \in [n] \setminus C\} \cup \{(k+1) - \text{atoms}\} \rangle$.
- Let $a \perp\!\!\!\perp b|C$ be any k-atom for $k < n - 2$. Then we have $\langle a \perp\!\!\!\perp b|C \cup \{(k-1) - \text{atoms}\} \rangle = \langle \{x \perp\!\!\!\perp y|Z : xyZ = abC\} \cup \{(k-1) - \text{atoms}\} \rangle$.

Direct Sums

Corollary. Any maximal semigraphoid that is not equal to $\langle \text{non} - k - \text{atoms} \rangle$ for any k, must have at least one l-atom for every $0 \leq l \leq n - 2$.

Let A and B be two disjoint sets of random variables. Given two semigraphoids S, T such that $S \subseteq CI(A)$ and $T \subseteq CI(B)$, we define the direct sum [3] of S and T to be the semigraphoid:

$$S \oplus T = \langle S \cup T \cup \{A \perp\!\!\!\perp B\} \rangle.$$

The direct sum operation is associative. For any semigraphoids $\mathcal{S}$, $\mathcal{T}$, and $\mathcal{U}$ on disjoint sets we have

$$\mathcal{S} \oplus (\mathcal{T} \oplus \mathcal{U}) = (\mathcal{S} \oplus \mathcal{T}) \oplus \mathcal{U}.$$

Lemma. Let N be an index set and $A \subset N$. The direct sum of the complete model on A and the complete model on $N \setminus A$ is $CI(N)$.

Theorem. Let N be an index set and $A \subset N$. The direct sum of a maximal semigraphoid on A and the complete model on $N \setminus A$ is a maximal semigraphoid on N.

The following direct sums are maximal on four random variables:

1. $\langle a \perp\!\!\!\perp bc \rangle \oplus \langle \varnothing \rangle_{\{d\}} = \langle \varnothing \rangle_{\{a\}} \oplus \langle \varnothing \rangle_{\{b,c\}} \oplus \langle \varnothing \rangle_{\{d\}} = \langle a \perp\!\!\!\perp d \rangle \oplus \langle \varnothing \rangle_{\{b,c\}}$
2. $\langle a \perp\!\!\!\perp b, a \perp\!\!\!\perp c, b \perp\!\!\!\perp c \rangle \oplus \langle \varnothing \rangle_{\{d\}}$
3. $\langle a \perp\!\!\!\perp b|c, a \perp\!\!\!\perp c|b, b \perp\!\!\!\perp c|a \rangle \oplus \langle \varnothing \rangle_{\{d\}}$

Maximal on Four Random Variables

All maximal semigraphoids on four random variables up to permutation (for arbitrary $a, b, c, d \in [4]$):

For $k \in \{0, 1, 2\}$, the first three maximal semigraphoids are the non-k-atoms.

4. $\langle \cup_{k \in \{0,1\}} \{k - \text{atoms}\} \rangle$
5. $\langle \cup_{k \in \{0,2\}} \{k - \text{atoms}\} \rangle$
6. $\langle \cup_{k \in \{1,2\}} \{k - \text{atoms}\} \rangle$

The last four follow definite patterns:

7. $\langle CI(3) \cup \{2 - \text{atoms}\} \rangle = \langle \begin{bmatrix} a \perp\!\!\!\perp b \\ a \perp\!\!\!\perp c \\ b \perp\!\!\!\perp c \end{bmatrix}, \begin{bmatrix} a \perp\!\!\!\perp b|c \\ a \perp\!\!\!\perp c|b \\ b \perp\!\!\!\perp c|a \end{bmatrix}, \{2 - \text{atoms}\} \rangle$
8. $\langle \{0 - \text{atoms}\}, \begin{bmatrix} a \perp\!\!\!\perp b|c \\ a \perp\!\!\!\perp c|b \\ b \perp\!\!\!\perp c|a \end{bmatrix}, \begin{bmatrix} a \perp\!\!\!\perp b|cd \\ a \perp\!\!\!\perp c|bd \\ b \perp\!\!\!\perp c|ad \end{bmatrix} \rangle$

Observe that these last two are formed by strategically choosing a *chain* of atoms without consequences.

	0-atoms	1-atoms	2-atoms
9	$\langle\ a \perp\!\!\!\perp b,$	$\sim \begin{bmatrix} a \perp\!\!\!\perp x\|b \\ b \perp\!\!\!\perp x\|a \end{bmatrix}_{x \in [4] \setminus ab},$	$c \perp\!\!\!\perp d\|ab\ \rangle$
10	$\langle\ \sim a \perp\!\!\!\perp b,$	$\begin{bmatrix} a \perp\!\!\!\perp x\|b \\ b \perp\!\!\!\perp x\|a \end{bmatrix}_{x \in [4] \setminus ab},$	$\sim c \perp\!\!\!\perp d\|ab\ \rangle$

Algorithm Sketch

An algorithm for finding all maximal semigraphoids on n random variables begins with the non-k-atoms. For every set of non-k-atoms, we add a k-atom to the set while simultaneously choosing either $(k-1)$- or $(k+1)$-atoms to remove that *interact* with that k-atom (have consequences). Unless our maximal semigraphoid is a chain of atoms that do not interact with one another, they will be composed of sets of k-atoms for every k except two consecutive, say l and $l+1$.

Outlook

- A *graphoid* $\mathcal{G}$ is a semigraphoid that satisfies a fifth axiom for every pairwise disjoint $A, B, C, D \subset [n]$: Intersection: If $\{A \perp\!\!\!\perp B|CD, A \perp\!\!\!\perp C|BD\} \subset \mathcal{G}$, then $A \perp\!\!\!\perp BC|D \in \mathcal{G}$. [1]
- As graphoids are semigraphoids by definition, we extend our study to these structures. Out of the ten types of maximal semigraphoids classified above, the semigraphoids of type 1,2, and 4 are all of the maximal graphoids.
- Semigraphoids on n discrete random variables define algebraic varieties [1,2]. We plan to further investigate the conditional independence ideals of these maximal sets. Addressing computational complications that the contraction axiom presents.

References

[1] M. Drton,B. Sturmfels,S. Sullivant, Lectures on Algebraic Statistics, Oberwolfach Seminars, **39**, Basel, 2009.

[2] L. D. Garcia, M. Stillman, B. Sturmfels, *Algebraic Geometry of Bayesian Networks*, Journal of Symbolic Computation, **39**, pp. 331-355, 2005.

[3] F. Matus, *Towards Classification of Semigraphoids*, Discrete Mathematics, **277**, pp. 115-145, 2004.

[4] M. Studeny, *Probabilistic Conditional Independence Structures*, Springer Series in Information Science and Statistics, Springer-Verlag, London, UK, 2005.

SAN DIEGO STATE UNIVERSITY