Graphs of random variables

A *graph* is a pair $\mathcal{G} = (V, E)$ where

- V is a set of *vertices* or *nodes*
- $E \subseteq V \times V$ with $(v, v) \notin E$ for any $v \in V$ is a set of *edges*.

Let $Z = (Z_1, \ldots, Z_p)^T$ be a collection of random variables. The graphs we will consider will have $V = \{1, \ldots, p\}$.

Graph terminology

Let $j, k \in V$.

- We say there is an *edge* between j and k and that j and k are *adjacent* if either $(j, k) \in E$ or $(k, j) \in E$.
- An edge (j, k) is *undirected* if also $(k, j) \in E$; otherwise it is *directed* and we may write $j \to k$ to represent this.
- If all edges in the graph are (un)directed we call it an *(un)directed graph*.

Graph terminology

- A graph $\mathcal{G}_1 = (V_1, E_1)$ is a *subgraph* of $\mathcal{G} = (V, E)$ if $V_1 \subseteq V$ and $E_1 \subseteq E$ and a *proper subgraph* if either of these are proper inclusions.
- Say j is a *parent* of k and k is a *child* of j if $j \to k$. The sets of parents and children of k will be denoted $\mathrm{pa}(k)$ and $\mathrm{ch}(k)$ respectively.
- A set of three nodes is called a *v-structure* if one node is a child of the two other nodes, and these two nodes are not adjacent.
- The *skeleton* of $\mathcal{G}$ is a copy of $\mathcal{G}$ with every edge replaced by an undirected edge.

Graph terminology

- A *path* from j to k is a sequence $j = j_1, j_2, \ldots, j_m = k$ of (at least two) distinct vertices such that j_l and j_{l+1} are adjacent.
- It is a *directed path* if $j_l \to j_{l+1}$ for all l. We then call k a *descendant* of j. The set of descendants of j will be denoted $\mathrm{de}(j)$.
- If $j_{l-1} \to j_l \leftarrow j_{l+1}$, j_l is called a *collider (relative to the path)*.
- A *directed cycle* is (almost) a directed path but with the start and end points the same. A *partially directed acyclic graph (PDAG)* is a graph containing no directed cycles.
- A *directed acyclic graph (DAG)* is a directed graph containing no directed cycles.

Graph terminology

- In a DAG, a path between j_1 and j_m $(j_1, j_2, \ldots, j_m)$ is *blocked by a set (of nodes) S* with neither j_1 nor j_m in S whenever there is a node j_l such that one of the following two possibilities hold:
  1. $j_l \in S$ and we *don't* have $j_{l-1} \to j_l \leftarrow j_{l+1}$
  2. $j_{l-1} \to j_l \leftarrow j_{l+1}$ and neither j_l nor any of its descendants are in S.
- If $\mathcal{G}$ is a DAG, given a triple of (disjoint) subsets of nodes A, B, S, we say S *d-separates* A from B if S blocks every path from A to B.

Graph terminology

- Given a triple of (disjoint) subsets of nodes A, B, S, we say S *separates* A from B if every path from a node in A to a node in B contains a node in S.
- The *moralised graph* of a DAG $\mathcal{G}$ is the undirected graph obtained by adding edges between (marrying) the parents of each node and removing all edge directions.