Introduction to Directed Graphs

Topics Covered: Vertices, Edges – types of, D-separation,

Motivation: Oftentimes we are uncertain about which variable causes which variable in a modeling effort. Theory may tell us what our fundamental causal variables are in a controlled system; however, oftentimes our data are not collected in a controlled environment. In fact we are rarely involved with the collection of our data.

So our data are said to be observational (non-experimental) and usually secondary, not collected explicitly for our purpose but rather for some other primary purpose.

Theory usually invokes the *ceteris paribus* condition to achieve results. Data are usually observational (non-experimental) and thus the *ceteris paribus* condition may not hold. We may not ever know if it holds because of unknown variables operating on our system. If we know the true system and know our data are appropriate measures of the variables in that true system, then multiple regression techniques are appropriate. If we do not know the "true" system, but have an approximate idea that one or more variables operate on that system then experimental methods can yield appropriate results, because they use randomization (random assignment of subjects to alternative treatments) to account for any additional variation associated with the unknown variables of the system.
Directed graphs help us assign causal flows to a set of data which the problem (theory) suggests should be related, even if we do not know the "true" system.

Directed graphs work off of the variance-covariance matrix from a set of variables. We will learn one algorithm for building such models in the next few lectures.

Recently Papineau (1985) has discussed a non-time induced asymmetry in causal relations – causal forks and inverted causal forks.

Causal Forks

For variables X, Y and Z, we say X causes Y and Z, illustrated as: Y \leftarrow X \rightarrow Z, if the unconditional association between Y and Z is nonzero, but the conditional association between Y and Z given knowledge of the common cause X, is zero. Common causes screen off associations between their joint effects. Common causes induce a causal fork.

Example of a Causal Fork

Consider the three events (variables): Y= Bill has yellow stain between his index and middle fingers; Z = Bill has lung cancer; X= Bill has smoked for forty years. Here we have a causal fork: X causes Y and X causes Z.

\[ X \rightarrow Y \quad X \rightarrow Z \]
Now if we do not observe X, but observe Y and Z, we might notice that a lot of people with yellow stained fingers have cancer. But we understand that the yellow stain didn’t cause their cancer. Further, we understand that washing their hands with a yellow stain solvent soap won’t cure their cancer. This is recognition, in a very simple way the adage that correlation does not imply causation. Once we observe the fact that Bill smokes (we observe X) then the extra information that he has yellow fingers contributes nothing to our knowledge of his having cancer.

**Inverted Causal Fork**

On the other hand, we say X and Z cause Y, illustrated as:

\[ X \rightarrow Y \leftarrow Z, \]

if the unconditional association between X and Z is zero, but the conditional association between X and Z given the common effect Y is not zero. *Common effects do not screen off association between their joint causes*. This is an inverted causal fork.

**Example of an Inverted Causal Fork**

Think of common effects:
X is my battery is dead.
Z is my gasoline tank is empty.
Y is the event that my car won’t start.
Now X and Z are generally not related (give me some literary licence here); but if I go outside in the morning to start my car and note (with some surprise and anger) that it won’t start, I immediately think of two causes (X and Z). Now if I go back into my house and ask my son if he remembered to fill the tank before he got home last night and he responds “no and the car sputtered into the driveway when I got home,” my mind immediately assignes a high probability to Z (I’m out of gas) and a small probability to X (my battery is dead).

So knowledge of Y (my car won’t start) and Z (I’m out of gas) gives me some knowledge of X (my battery probably is ok).

Papineau’s perception has been built upon over the last decade plus giving us a rather expansive literature, reviewed here under the heading of directed graphs.

A graph is a picture, which can be represented as ordered triple <V,M,E> where V is a non-empty set of vertices (variables), M is a non-empty set of marks (symbols attached to the end of undirected edges), and E is a set of ordered pairs.
Each member of $E$ is called an edge (line).

Vertices (variables) connected by an edge are said to be adjacent. If we have a set of vertices \{A, B, C, D, E\}:

(i) the undirected graph contains only undirected edges: (e.g., $A \rightarrow B$);

(ii) a directed graph contains only directed edges (e.g., $B \rightarrow C$);

(iii) an inducing path graph contains both directed edges and bi-directed edges (e.g., $C \leftrightarrow D$);

(iv) a partially oriented inducing path graph contains directed edges ($\rightarrow$), bi-directed edges ($\leftrightarrow$), non-directed edges ($\rightarrow\rightarrow$) and partially directed edges ($\rightarrow\rightarrow$).

A directed acyclic graph is a directed graph that contains no directed cyclic paths (an acyclic graph contains no vertex more than once). Only acyclic graphs are used in this course.

We do not allow cyclic graphs (edges lead away from a variable and then back to the same variable):

\[
\begin{array}{c}
\text{A} \\
\mathrm{B}
\end{array}
\]
Directed acyclic graphs are designs for representing conditional independence as implied by the recursive product decomposition:

\[
\text{(1)} \quad \Pr(x_1, x_2, x_3, \ldots x_n) = \prod_{i=1}^{n} \Pr(x_i \mid \text{pa}_i),
\]

where \( \Pr \) is the probability of vertices \( x_1, x_2, x_3, \ldots x_n \) and \( \text{pa}_i \) the realization of some subset of the variables that precede (come before in a causal sense) \( X_i \) in order \( (X_1, X_2, \ldots, X_n) \).

Pearl (1995) proposes d-separation as a graphical characterization of conditional independence. That is, d-separation characterizes the conditional independence relations given by equation (1).

If we formulate a directed acyclic graph in which the variables corresponding to \( \text{pa}_i \) are represented as the parents (direct causes) of \( X_i \), then the independencies implied by equation (1) can be read off the graph using the notion of d-separation (defined in Pearl (1995)):

**Definition:** Let \( X, Y \) and \( Z \) be three disjoint subsets of vertices in a directed acyclic graph \( G \), and let \( p \) be any path between a vertices in \( X \) and a vertices in \( Y \), where by 'path' we mean any succession of edges, regardless of their directions. \( Z \) is said to block \( p \) if there is a vertex \( w \) on \( p \) satisfying one of the following: (i) \( w \) has converging arrows along \( p \), and neither \( w \) nor any of its descendants are on \( Z \), or, (ii) \( w \) does not have converging arrows along \( p \), and \( w \) is in \( Z \). Further, \( Z \) is said to d-separate \( X \) from \( Y \) on graph \( G \), written \( (X \perp Y \mid Z)_G \), if and only if \( Z \) blocks every path from a vertex in \( X \) to a vertex in \( Y \).
Geiger, Verma and Pearl (1990) show that there is a one-to-one correspondence between the set of conditional independencies, $X \perp Y \mid Z$, implied by equation (1) and the set of triples $(X, Y, Z)$ that satisfy the d-separation criterion in graph $G$.

**KEY TO UNDERSTANDING THESE IDEAS IS THAT D-SEPARATION ALLOWS US TO WRITE THE PROBABILITY OF OUR VARIABLES $X, Y, X$ IN TERMS OF THE PRODUCT OF THE CONDITIONAL PROBABILITY ON EACH VARIABLE ($X, Y, \text{ OR } Z$) WHERE THE CONDITIONING FACTOR IS JUST THE IMMEDIATE PARENT OF EACH VARIABLE. WE DO NOT HAVE TO CONDITION ON GRANDPARENTS OR GREAT GRANDPARENTS!!!** SEE BELOW:

\[
\begin{align*}
A \rightarrow B \leftarrow C & ; \quad P(ABC) = P(A)P(C)P(B|CA) \\
D \rightarrow E \rightarrow F & ; \quad P(DEF) = P(D)P(E|D)P(F|E) \\
G \rightarrow H \rightarrow I \leftarrow J & ; \quad P(GHIJ) = P(G)P(J)P(H|G)P(I|H,J) \\
K \rightarrow L \rightarrow M \rightarrow N & \\
& \quad \uparrow \\
& \quad O ; \quad P(KLNO) = P(K)P(L|K)P(O)P(M|L,O)P(N|M)
\end{align*}
\]

Here we are looking at traffic fatalities and want to see what variables "cause" them in directed graphs -- we took a model from another researcher who said all of the variables listed below should be the cause of traffic fatalities. We get a slightly different picture. Next time we'll study how to build these graphs (\(t\) means change: period t level minus period t-1 level).

The variables under study were:

- traffic fatalities
- speed limit
- young drivers
- mileage driven
- income
- cost of safety devices
- alcohol consumption
- trend
We study these in differences from year to year and find the following:

Figure 2. Causally Sufficient Graphical Replication of Peltzman’s Model for 1947-1965 Differenced Data Using A 20% Significance Level.

This graph suggests that income and mileage are related but it is income and not mileage that cause traffic fatalities. My guess is that if we did an experiment we would find just the opposite result. But I believe we probably measure income better than we measure mileage, so, because we are using observational (not experimental) data we get results that need some interpretation.

The result on young drivers is interesting (age 15 – 26). I suspect that if we used an age range of 15 – 17 we would find an edge between youth and traffic fatalities.