14.3 The operation of arc reversal in a Bayesian network allows us to change the direction of an arc X → Y while preserving the joint probability distribution that the network represents (Shachter, 1986). Arc reversal may require introducing new arcs: all the parents of X also become parents of Y, and all parents of Y also become parents of X.
a. Assume that X and Y start with m and n parents, respectively, and that all variables have k values. By calculating the change in size for the CPTs of X and Y, show that the total number of parameters in the network cannot decrease during arc reversal. (Hint: the parents of X and Y need not be disjoint.)
b. Under what circumstances can the total number remain constant?
c. Let the parents of X be U ∪ V and the parents of Y be V ∪ W, where U and W are disjoint. The formulas for the new CPTs after arc reversal are as follows:

$$P(Y \mid U, V, W) = \sum_x P(Y \mid V, W, x) \times P(x \mid U, V)$$
$$P(X \mid U, V, W, Y) = P(Y \mid X, V, W) \times P(X \mid U, V) / P(Y \mid U, V, W)$$

Prove that the new network expresses the same joint distribution over all variables as the original network.

a. Assume that X and Y start with m and n parents, respectively, and that all variables have k values. By calculating the change in size for the CPTs of X and Y, show that the total number of parameters in the network cannot decrease during arc reversal. (Hint: the parents of X and Y need not be disjoint.)

Paper: Evaluating Influence Diagrams
Author: Ross D. Shachter
Year: 1986
http://cs.ru.nl/~peterl/BN/shachter1987.pdf

An influence diagram is a generalization of a Bayesian network. For this exercise we are interested in the Theorem 3 of the paper that says:

Theorem 3: Arc Reversal. *Given that there is an arc (i, j) between chance nodes i and j, but no other directed (i, j)-patht in a regular influence diagram, arc (i, j) can be replaced by arc (j, i). Afterward, both nodes inherit each other's conditional predecessors.*

Lets call $\#parents(X_{BN1})$ the number of parents that the variable X presents in the Bayesian Network 1 (before arc reversal) and $\#parents(X_{BN2})$ the number of parents that the variable X presents in the Bayesian Network 2 (after reversal).

$$\#parents(X_{BN1}) = m$$
$$\#parents(Y_{BN1}) = n$$

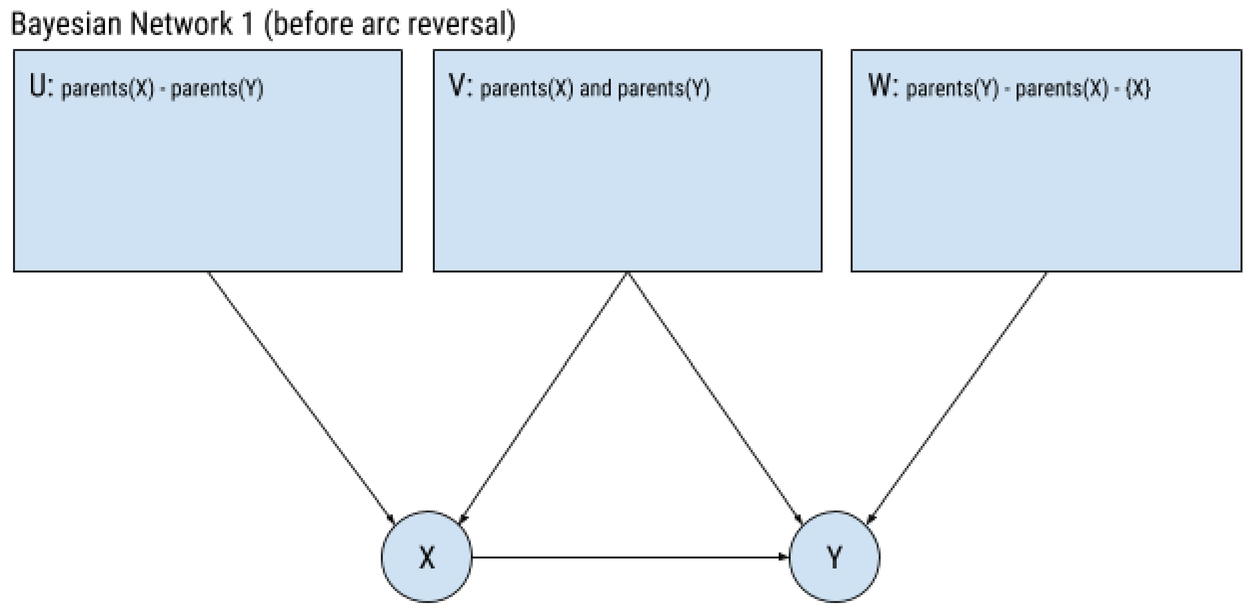
The theorem 3 says that after reversal the variable X will inherit all the parents of Y and Y will inherit all the parents of X. Then after the arc reversal we will find that:

$$parents(X_{BN1}) \subseteq parents(Y_{BN2})$$
$$parents(Y_{BN1}) \subseteq parents(X_{BN2})$$

The above conditions mean that:

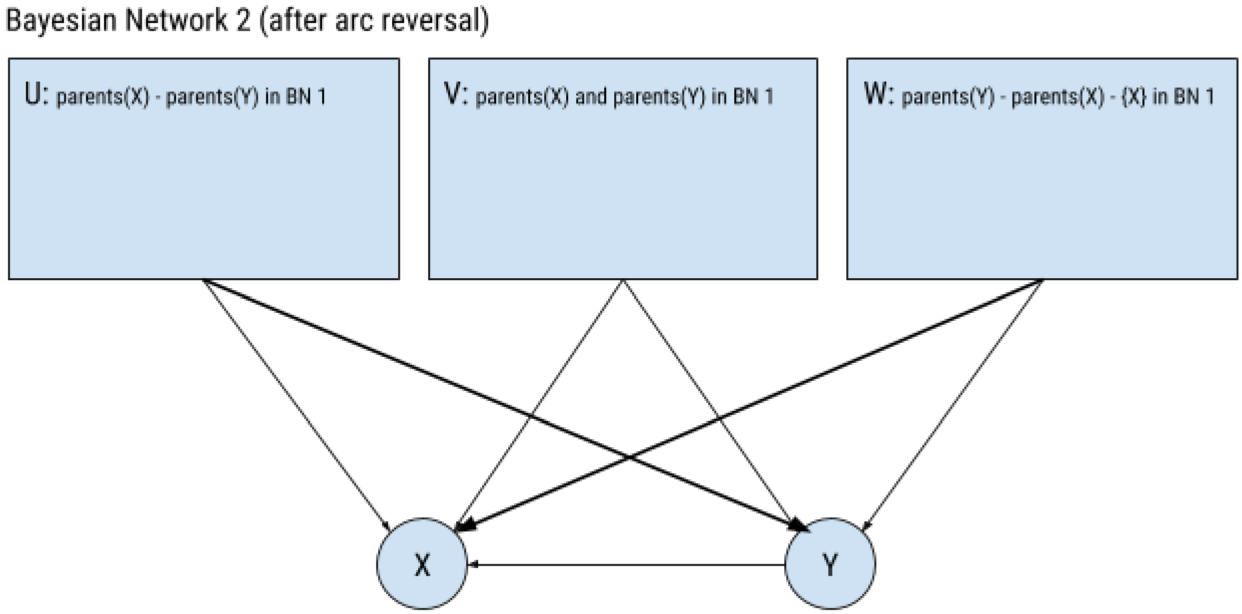
$\#parents(X_{BN2}) \geq \#parents(Y_{BN1}) = n$
$\#parents(Y_{BN2}) \geq \#parents(X_{BN1}) = m$
$\#parents(X_{BN1}) + \#parents(Y_{BN1}) = m + n$

And we can generalize the Bayesian Network as:



That let us rewrite m and n as:

$\#parents(X_{BN1}) = m = u + v$
$\#parents(Y_{BN1}) = n = w + v + 1$



$\#parents(X_{BN2}) = u + v + w + 1 = n + u$
$\#parents(Y_{BN2}) = u + v + w = m + w$

$\#parents(X_{BN2}) + \#parents(Y_{BN2}) = n + u + m + w$

When all variables of a Bayesian Network presents the same number of possible discrete values k, the size of the CPT of each variable can be easily calculated knowing the number of parents links of the variable.

For a simple variable we will need to save $k-1$ probability values. This is because the last value can be calculated based on the Axiom 1 of Probability Theory: probabilities of all atomic events must sum 1.

$size(CPT(X_{BN1})) = k^{parents(X_{BN1})} \times (k-1) = k^m \times (k-1)$
$size(CPT(Y_{BN1})) = k^{parents(Y_{BN1})} \times (k-1) = k^n \times (k-1)$

$\#paraments(X_{BN1} + Y_{BN1}) = size(CPT(X_{BN1})) + size(CPT(Y_{BN1})) = (k^m + k^n) \times (k-1)$

$size(CPT(X_{BN2})) = k^{parents(X_{BN2})} \times (k-1) = k^{n+u} \times (k-1)$
$size(CPT(Y_{BN2})) = k^{parents(Y_{BN2})} \times (k-1) = k^{m+w} \times (k-1)$

$\#paraments(X_{BN2} + Y_{BN2}) = size(CPT(X_{BN2})) + size(CPT(Y_{BN2})) = (k^{n+u} + k^{m+w}) \times (k-1)$

$ChangeRatio = \frac{\#paraments(X_{BN2}+Y_{BN2})}{\#paraments(X_{BN1}+Y_{BN1})} = \frac{(k^{n+u}+k^{m+w})\times(k-1)}{(k^n+k^m)\times(k-1)}$

$ChangeRatio = \frac{(k^{n+u}+k^{m+w})}{(k^n+k^m)}$

The number of parameters required can not decrease after arc reversal.

Q.E.D.

b. Under what circumstances can the total number remain constant?

The total number of parameters remains constant when u and w are zero (when U and W are empty groups). This is when there are no variables parents of X and not parents of Y and vice versa. Observing at the second Bayesian Network Graph after arc reversal, there are one link that change direction (from X->Y to Y->X) and there are new groups of links (arcs) that appears that are originated when U and W groups are not empties.

c. Let the parents of X be U ∪ V and the parents of Y be V ∪ W, where U and W are disjoint. The formulas for the new CPTs after arc reversal are as follows:

$$P_{BN2}(Y \mid U, V, W) = \sum_x P_{BN1}(Y \mid V, W, x) \times P_{BN1}(x \mid U, V)$$
$$P_{BN2}(X \mid U, V, W, Y) = P_{BN1}(Y \mid X, V, W) \times P_{BN1}(X \mid U, V) / P_{BN1}(Y \mid U, V, W)$$

Prove that the new network expresses the same joint distribution over all variables as the original network.

The definition of U, V and W here are the same I used to proof the items a) and b) of the exercise.

U : group of nodes that are parents of X and not parent of Y
V : group of nodes that are parents of X and parent of Y at the same time
W : group of nodes that are parents of Y and not parent of X with exception of X.

Lets complete the Bayesian Network 1 Universe:

D : group of nodes that are descendants of X and Y and does not belong to $U \cup V \cup W \cup \{X, Y\}$
$\overline{D}$: group of nodes complement of D and does not belong to $U \cup V \cup W \cup \{X, Y\}$

Now every possible variable of the Bayesian Network must be located into one of the defined groups that build the entire universe: $U \cup V \cup W \cup \{X\} \cup \{Y\} \cup D \cup \overline{D}$

This definitions of the groups is very important because they are independent groups that will help to estimate some probabilities in a while.

The arc reversal process only makes arc modifications between nodes in groups U, V, W and {X, Y}. The arc reversal process does not add or removes arcs in nodes that belong to the groups $D \text{ or } \overline{D}$.

Remember Equation 14.2 (AIMA 3E, p.513) to calculate a generic entry in the joint distribution:

$$P(x_1, x_2, \ldots, x_n) = \prod_{i=1}^{n} P(x_i \mid parents(X_i))$$

Then the full joint distribution before arc removal could be defined as $P(\overline{D}, U, V, W, X, Y, D)$:

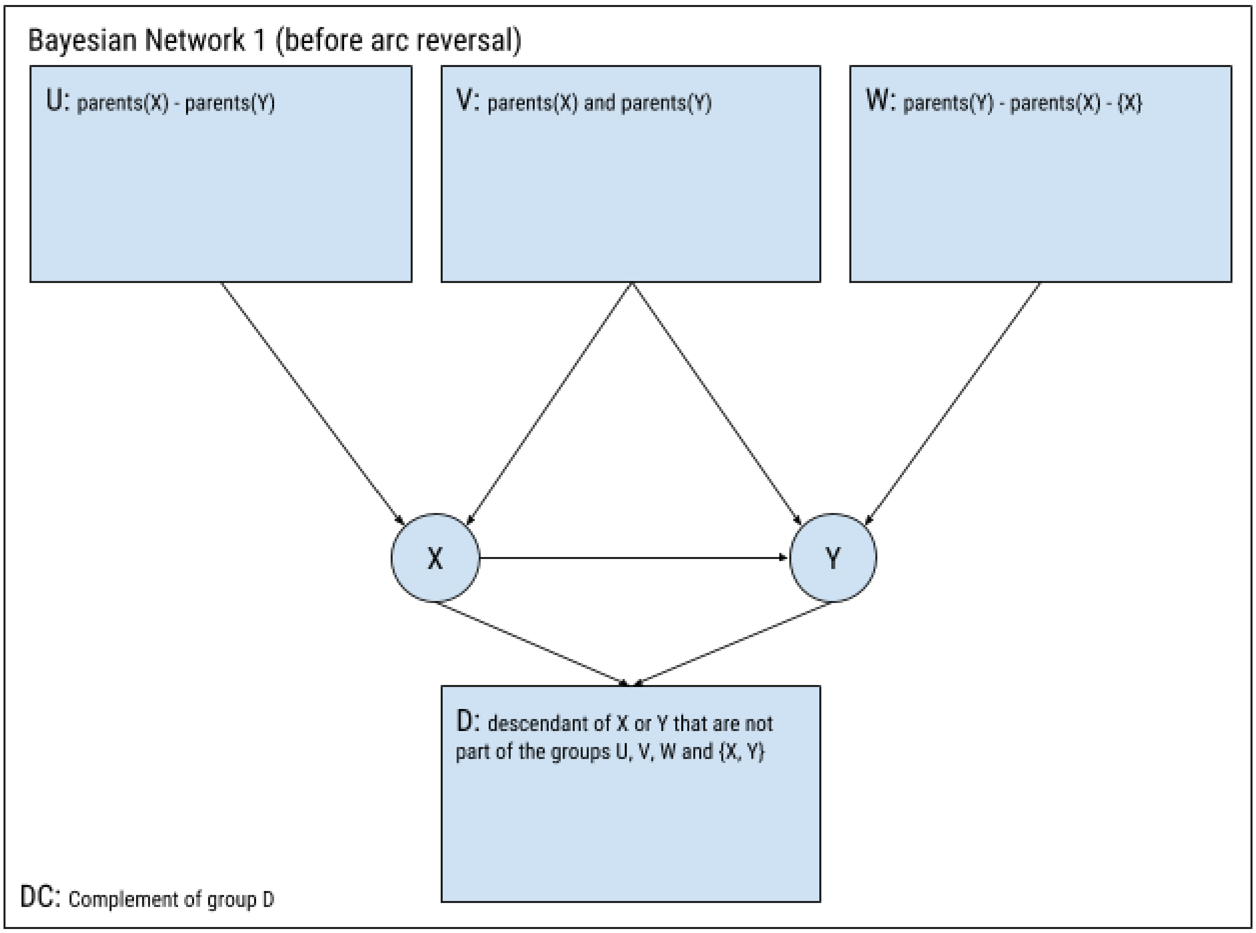
$$P(\overline{D}, U, V, W, X, Y, D) = P((\overline{D}, U, V, W), (X, Y), (D))$$

I) By Equation 14.2:

$$P_{BN1}(\overline{D}, U, V, W, X, Y, D) = P_{BN1}(\overline{D}, U, V, W) \times P_{BN1}(X, Y \mid U, V, W) \times P_{BN1}(D \mid U, V, W, X, Y)$$

So, we could prove the new network expresses the same joint distribution over all variables as the original network if we prove that:

a) $P_{BN1}(\overline{D}, U, V, W) = P_{BN2}(\overline{D}, U, V, W)$
b) $P_{BN1}(X, Y \mid U, V, W) = P_{BN2}(X, Y \mid U, V, W)$
c) $P_{BN1}(D \mid U, V, W, X, Y) = P_{BN2}(D \mid U, V, W, X, Y)$



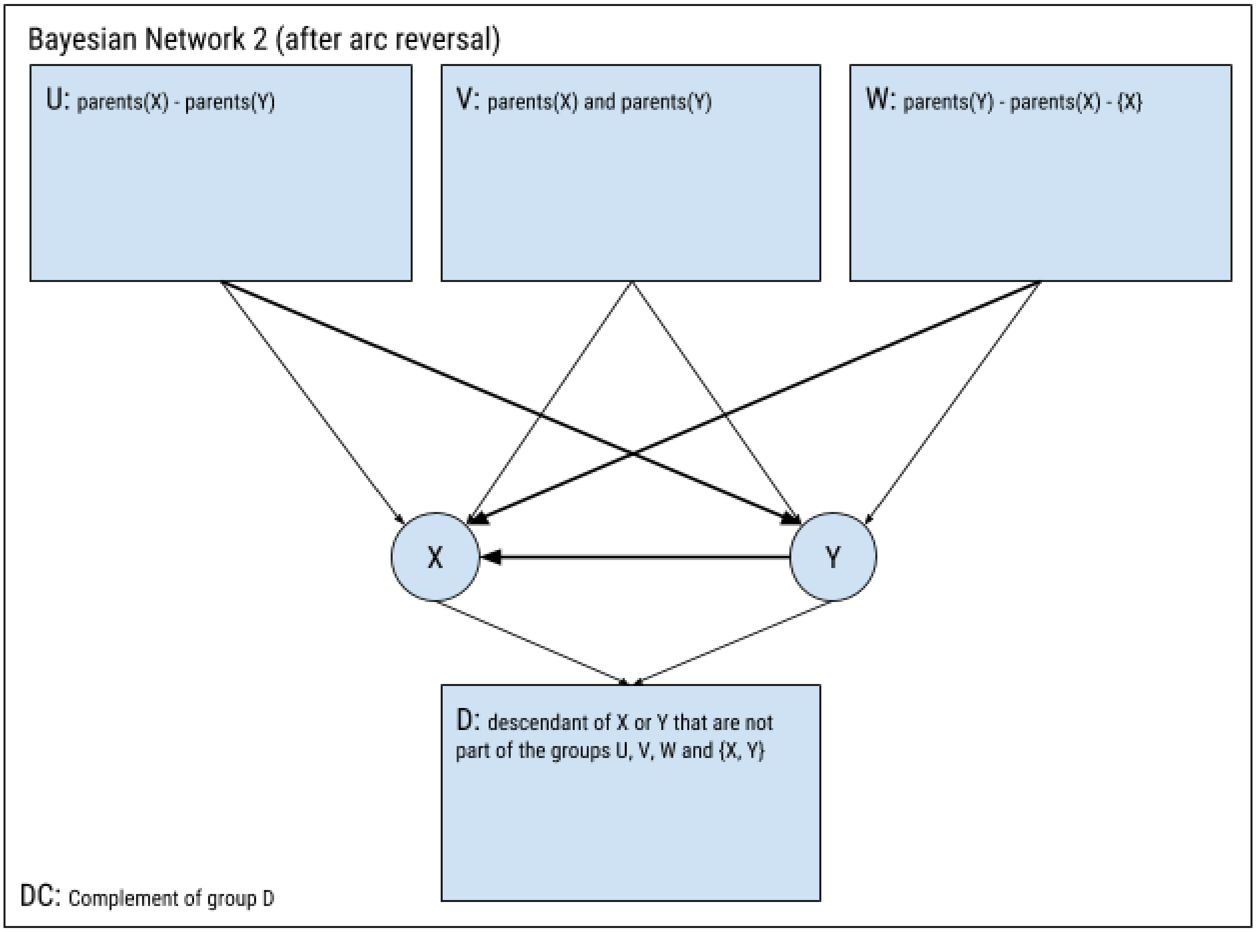
Proof of (a) and (c):

The arc reversal process does not change CPT of the variables that belongs to groups $(\overline{D}, U, V, W)$, then the probabilities of the original network will be the same in the new network:

$$P_{BN1}(\overline{D}, U, V, W) = P_{BN2}(\overline{D}, U, V, W)$$

And by definition of the group D, all the variables in D depends completely of its parents and we know they are located in: $U, V, W,\ X,\ and\ Y$:

$$P_{BN1}(D \mid U, V, W, X, Y) = P_{BN2}(D \mid U, V, W, X, Y)$$



Proof of (b):

II) After arc reversal process we the exercise says that:

$$P_{BN2}(Y \mid U, V, W) = \sum_x P_{BN1}(Y \mid V, W, x) \times P_{BN1}(x \mid U, V)$$
$$P_{BN2}(X \mid U, V, W, Y) = P_{BN1}(Y \mid X, V, W) \times P_{BN1}(X \mid U, V) / P_{BN1}(Y \mid U, V, W)$$

III) By equation 14.2

$$P_{BN2}(X, Y \mid U, V, W) = P_{BN2}(Y \mid U, V, W) \times P_{BN2}(X \mid U, V, W, Y)$$

IV) Using equations in (II) in (III)

$$P_{BN2}(X, Y \mid U, V, W) = \frac{\sum_x P_{BN1}(Y \mid V, W, x) \times P_{BN1}(x \mid U, V) \times P_{BN1}(Y \mid X, V, W) \times P_{BN1}(X \mid U, V)}{P_{BN1}(Y \mid U, V, W)}$$

We can add to the first term the group U that is independent of variable Y before the arc reversal process.

$$= \frac{\sum_x P_{BN1}(Y \mid U, V, W, x) \times P_{BN1}(x \mid U, V) \times P_{BN1}(Y \mid X, V, W) \times P_{BN1}(X \mid U, V)}{P_{BN1}(Y \mid U, V, W)}$$

$$= \frac{\sum_x P_{BN1}(Y, U, V, W, x) \times P_{BN1}(U, V, W) \times P_{BN1}(x \mid U, V) \times P_{BN1}(Y \mid X, V, W) \times P_{BN1}(X \mid U, V)}{P_{BN1}(U, V, W, x) \times P_{BN1}(Y, U, V, W)}$$

$$= \frac{\sum_x P_{BN1}(x \mid Y, U, V, W) \times P_{BN1}(x \mid U, V) \times P_{BN1}(Y \mid X, V, W) \times P_{BN1}(X \mid U, V)}{P_{BN1}(x \mid U, V, W)}$$

And we know that variable x is independent of W, then:

$$= \frac{\sum_x P_{BN1}(x \mid Y, U, V, W) \times P_{BN1}(x \mid U, V) \times P_{BN1}(Y \mid X, V, W) \times P_{BN1}(X \mid U, V)}{P_{BN1}(x \mid U, V)}$$

$$P_{BN2}(X, Y \mid U, V, W) = \sum_x P_{BN1}(x \mid Y, U, V, W) \times P_{BN1}(Y \mid X, V, W) \times P_{BN1}(X \mid U, V)$$

$$P_{BN2}(X, Y \mid U, V, W) = 1 \times P_{BN1}(Y \mid X, V, W) \times P_{BN1}(X \mid U, V)$$

$$P_{BN2}(X, Y \mid U, V, W) = P_{BN1}(Y \mid X, V, W) \times P_{BN1}(X \mid U, V)$$

V) By equation 14.2

$$P_{BN1}(X, Y \mid U, V, W) = P_{BN1}(Y \mid X, V, W) \times P_{BN1}(X \mid U, V)$$

VI) By (IV) and (V)

$$P_{BN1}(X, Y \mid U, V, W) = P_{BN2}(X, Y \mid U, V, W)$$

Then we found a proof for:

d) $P_{BN1}(\overline{D}, U, V, W) = P_{BN2}(\overline{D}, U, V, W)$
e) $P_{BN1}(X, Y \mid U, V, W) = P_{BN2}(X, Y \mid U, V, W)$
f) $P_{BN1}(D \mid U, V, W, X, Y) = P_{BN2}(D \mid U, V, W, X, Y)$

VII) And we can conclude:

$$P_{BN2}(\overline{D}, U, V, W, X, Y, D) = P_{BN1}(\overline{D}, U, V, W) + P_{BN1}(X, Y \mid U, V, W) + P_{BN1}(D \mid U, V, W, X, Y)$$

VIII) That implies that:

$$P_{BN1}(\overline{D}, U, V, W, X, Y, D) = P_{BN2}(\overline{D}, U, V, W, X, Y, D)$$

Q.E.D.