Duplicate column filter FOR UNPACKING

Suppose A🡪B

(Note that there might be other things which point to B, and moreover there might be other things that A points to.)

Let

* U\_AB be all latent variables common to both A and B
* U\_A be all latent parents of A but not B
* U\_B be all latent parents of B but not A
* Pa\_B be all observable parents of B other than A (including private settings)

Take a given column.

If A=a for U\_AB=0 and **all** instances of U\_A:

then mark the column as a duplicate if for **any** {U\_B, Pa\_B} we find B\_(U\_AB=0,U\_B,A≠a, Pa\_B)>0.

Duplicate column filter FOR ANY INFLATION:

Suppose a column assigns all copies of **S** the same value, where **S** is some strict subset of the original variables.

* Let **U1/S** denote the subset of latent variables that have *at most one child outside of* **S**
* Let **U∞/S** denote the subset of latent variables whole children are *entirely* outside of **S**

Then, **if** the set of all latent variables is comprised entirely of **U1/S** union with **U∞/S:**

Let B be any single variable outside of **S**. Consider all copies of B in the inflation graph with **U∞/S=0** but **U1/S** varying. Define some “canonical ordering” for these copies of B (arbitrary, but should depend only on copy index, not value assignment).

If this column (which assigns all copies of **S** the same value) has an assignment of values to these copies of B such that the values assignment tuple is not ordered (relative to the arbitrary ordering of these copies of B), mark the column as duplicate.