Transitive closure in Record Matching.

Record1

Record2

Record3

If record1 matches record 2 but does not matches record3, and record2 matches record3 then eventually record1 is considered to be matching record3.

This is used while Data integration during Entity Resolution.

Consider the example:

ReferenceID, Firstname, LastName, SSN, DOB

Ref1, Abhi, Kumar, 123, 9/05/1991

Ref2, Abhi, Kumar, 1234, 9/05/1991

Ref3, Abhishek, Kumar, 1234, 09/05/1991

Ref4, Abhishek, Kr, 1234, 09/05/1991

Here we can see:

Based on firstname and Lastname Ref1, Ref2 matches,

Based on Lastname and DOB Ref1, ref2 matches and ref3, ref4 matches

Based on lastname and SSN Ref2, Ref3 matches

So the input of this program will be;

Ref1, Ref2

Ref1, Ref2

Ref3, Ref4

Ref2, Ref3

And the output will be:

Ref1, Ref2,Ref3, Ref4