Getting Rid of Unit Productions

Def 0.1 Let $G = (N, \Sigma, P, S)$ be a CFG. A production of the form $A \to B$ is a *unit production*.

NOTE- the algorithm below is a bit different than the one we did in class. But the ideas are similar.

Theorem 0.2 *There exists an algorithm that will, given a CFG $G = (N, \Sigma, P, S)$ with no useless nonterminals and no e-productions, will output a grammar $G' = (N', \Sigma, P', S')$ with no useless nonterminals, no e-production AND no unit productions such that $L(G) = L(G')$.*

Proof: We give the algorithm, show that it works in the correct time, but do not prove that it works.
We use $\Rightarrow$ to mean $\Rightarrow_G^*$. We first find all $A, B \in N$ such that $A \Rightarrow B$. Since there are no e-productions this is easy and only involves unit-productions. Formally we make a directed graph out of all of the nonterminals, with an edge between X and Y if $X \to Y$. Then, all pairs A, B such that there is a directed graph from A to B are all the pairs such that $A \Rightarrow B$.
Let the set of all (A, B) such that $A \Rightarrow B$ be called SUPERUNITS.

1. Let $PROD$ be P minus the UNIT productions.
2. Find the set of SUPERUNITES.
3. For all SUPERUNITS $A \Rightarrow B$ do
   (a) For all productions in $PROD$ of the form $X \to \alpha_1 A \alpha_2 A \cdots \alpha_{L-1} A \alpha_L$ add all productions that replace some of the A's with B's (there will be $2^L - 1$ new productions). Note that these new productions are now in $PROD$.

∎