1 Informal proof of the convergence of RankProp

Let $Y_U(t) = [y_2(t), ..., y_m(t)]^T$ and Y_U^* denote the limit of the sequence $\{Y_U(t)\}$. So $Y_U^* = [y_{s+1}^*, ..., y_m^*]^T$. Then we have $Y_U(t+1) = K_{U1}Y_1(0) + \alpha K_{UU}Y_U(t)$, where K_{UU} and K_{U1} are the sub-matrices of K with $U = \{2, ..., m\}$. Substituting the limit Y_U^* for $Y_U(t)$ and $Y_U(t+1)$, and writing Y_S for $Y_1(0)$, then $Y_U^* = K_{U1}Y_S + \alpha K_{UU}Y_U^*$, which can be transformed into $(I - \alpha K_{UU})Y_U^* = K_{U1}Y_S$. By definition of the normalization of K, the submatrix K_{UU} is stochastic, and hence its eigenvalues λ satisfy $|\lambda| \leq 1$. Therefore $I - \alpha K_{UU}$ is invertible for $0 \leq \alpha < 1$, and we have

$$Y_U^* = (I - \alpha K_{UU})^{-1} K_{U1} Y_S.$$

Thus, Y_U^* can be computed directly without iteration. However, such a calculation requires inverting a very large matrix, and it can be computationally more efficient simply to run a few iterations of the iterative scheme instead. In our experiments, we limit the number of iterations to 20, and we show empirically that this is enough to be close to convergence. Choosing the number of iterations and α are somewhat similar: if one requires a ranking with more local than global information in it, then one can either fix α large and perform few iterations, or fix α small and perform many iterations.