

SVD for word-doc matrix

$$A_{m \times n} = \begin{bmatrix} \text{US} \\ \text{UN} \\ \text{global} \\ \text{gene} \\ \text{lab} \end{bmatrix} \begin{bmatrix} d_1 & \dots & d_n \\ \text{(weight)} \\ \text{freq} \end{bmatrix}_{m \times n}$$

$$= \underbrace{\begin{bmatrix} | & & | \\ u_1 & \dots & u_m \\ | & & | \end{bmatrix}}_{\text{top } k} \begin{bmatrix} | & & | \\ \Sigma \\ | & & | \end{bmatrix}_{m \times m} \begin{bmatrix} | & & | \\ \text{top } k \\ \hline v_1^T \\ \vdots \\ v_n^T \\ \hline \end{bmatrix}_{n \times n}$$

truncated SVD

- u_i 's are document clusters

$$u_1 = \begin{bmatrix} 0.3 \\ 0.5 \\ 0.6 \\ 0.01 \\ 0.01 \end{bmatrix} \left. \vphantom{\begin{bmatrix} 0.3 \\ 0.5 \\ 0.6 \\ 0.01 \\ 0.01 \end{bmatrix}} \right\} \text{"political"}$$

$$u_2 = \begin{bmatrix} 0.01 \\ 0.05 \\ 0.02 \\ 0.5 \\ 0.3 \end{bmatrix} \left. \vphantom{\begin{bmatrix} 0.01 \\ 0.05 \\ 0.02 \\ 0.5 \\ 0.3 \end{bmatrix}} \right\} \text{"science"}$$

$$\phi(\text{US}) = [0.3 \ 0.01]$$

$$\phi(\text{gene}) = [0.01 \ 0.3]$$