## **Document Clustering Revisited**

Which cluster does d belong to?  $\rightarrow$  Which  $\theta_i$  was used to generate d?

$$\begin{aligned} &\textbf{d=x_1x_2...x_L} \quad \text{where } x_i \in \textbf{V} \\ &\text{cluster}(\textbf{d}) = \text{arg max}_i \ p(\theta_i \mid \textbf{d}) \\ &= \text{arg max}_i \ p(\textbf{d} \mid \theta_i) p(\theta_i) \\ &= \text{arg max}_i \ \prod_{j=1}^L p(x_j \mid \theta_i) p(\theta_i) \\ &= \text{arg max}_i \ \prod_{w \in \textbf{V}} p(w \mid \theta_i)^{c(w,d)} p(\theta_i) \end{aligned}$$

$$& = \text{arg max}_i \ \prod_{w \in \textbf{V}} p(w \mid \theta_i)^{c(w,d)} p(\theta_i)$$

$$& = \frac{p(\textbf{d} \mid \theta_i) p(\theta_i)}{p(\textbf{d})}$$

$$& = \frac{p(\textbf{d} \mid \theta_i) p(\theta_i)}{\sum_{j=1}^k p(\textbf{d} \mid \theta_j) p(\theta_j)}$$

