Document Clustering Revisited

Which cluster does d belong to? ➔ Which θ_i was used to generate d?

$d=x_1x_2\ldots x_L$ where $x_i \in V$

$$\text{cluster}(d) = \arg\max_i p(\theta_i \mid d)$$
$$= \arg\max_i p(d \mid \theta_i)p(\theta_i)$$
$$= \arg\max_i \prod_{j=1}^{L} p(x_j \mid \theta_i)p(\theta_i)$$
$$= \arg\max_i \prod_{w \in V} p(w \mid \theta_i)^{c(w,d)} p(\theta_i)$$

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