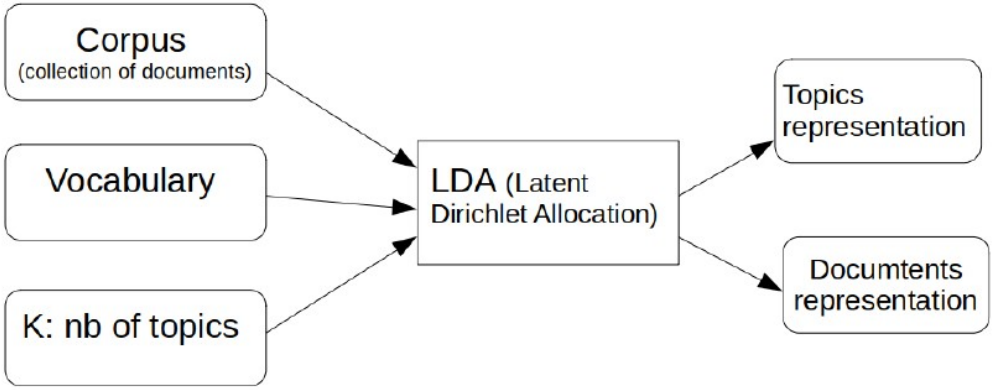
Topic models and the LDA algorithm



How could we use LDA decomposition to define documents similarities ?

β: Topics representation θ: Documents representation

$$\beta = \begin{bmatrix} \beta_{11} & \beta_{12} & \dots & \beta_{1V} \\ \vdots & \ddots & & \\ \beta_{K1} & \beta_{K2} & \dots & \beta_{KV} \end{bmatrix}$$

$$\theta = \begin{bmatrix} \theta_{11} & \theta_{12} & \dots & \theta_{1V} \\ \vdots & \ddots & & \\ \theta_{K1} & \theta_{K2} & \dots & \theta_{KV} \end{bmatrix}$$

K: Number of topics

V: Number of words

n: Number of documents

Distances between words
Word embeddings.

Two stages Wasserstein distance approach