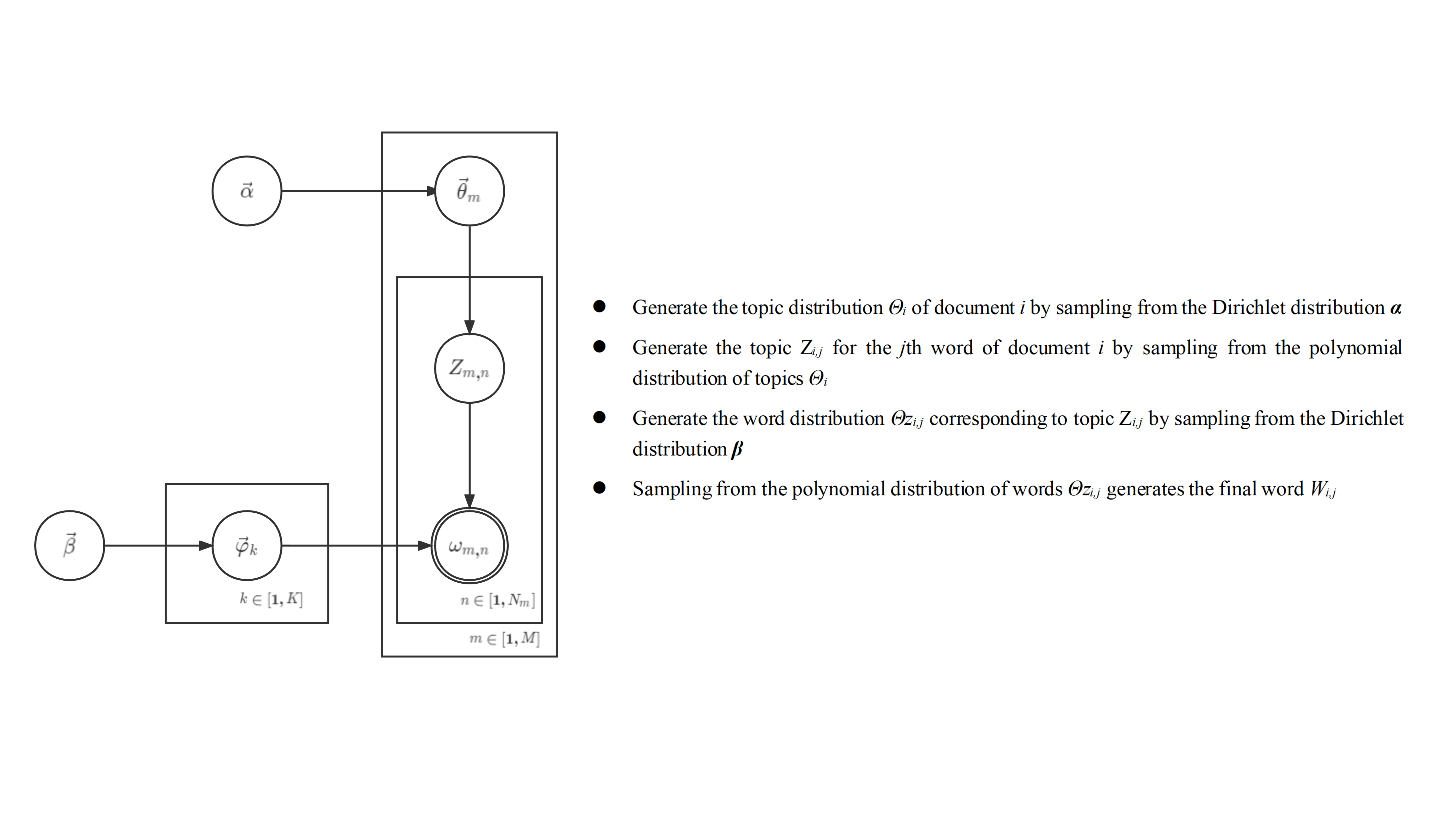
****

**Supplementary Figure 1. Graphical representation of the LDA model and document generation process.**

The analysis pipeline of the LDA model and the document generation process are shown in Supplementary Figure 1 (Blei, Ng, & Jordan, 2003). LDA is a three-level Bayesian probabilistic model and a document generation model with observable variables represented as double circles *w* and other latent variables represented as single circles. *α* is the Dirichlet parameter before each document topic distribution, *β* is the Dirichlet parameter before each topic word distribution, *Θⅰ* is the topic distribution of document *i* (sum of *Θⅰ* is 1.0), *φk* is the word distribution of topic *k*, Z*i,j* is the topic of the *j*th word in document *i*, and *Wi,j* is the specific word. Regarding the parameter settings of the LDA learning algorithm, in this study, *α* and *β* were set to 0.1 and 0.01 respectively, which are common settings in the literature (Hao, Zhang, Wang, & Gao, 2017).

Blei, D. M., Ng, A., & Jordan, M. I. (2003). Latent dirichlet allocation. *The Journal of Machine Learning Research.* doi: 10.1162/jmlr.2003.3.4-5.993

Hao, H., Zhang, K., Wang, W., & Gao, G. (2017). A tale of two countries: International comparison of online doctor reviews between China and the United States. *International Journal of Medical Informatics*, 99, 37-44. doi: 10.1016/j.ijmedinf.2016.12.007