

Table S3

Tuning Results for Different Model Specifications in the Latent Dirichlet Allocation

Number of topics	Alpha	Beta	Number of iterations	Optimization interval	Topic coherence	Training time
15	0.01	0.01	1000	OFF	-1.675	09 min 19 sec
15	-	0.03	1000	ON-100	-2.410	11 min 12 sec
30	0.01	0.01	1000	OFF	-1.661*	15 min 33 sec
30	-	0.02	1000	ON-100	-2.707	09 min 00 sec
50	0.01	0.01	1000	OFF	-1.691	21 min 00 sec
50	-	0.01	1000	ON-100	-2.934	09 min 46 sec
60	0.01	0.01	1000	OFF	-1.702	19 min 31 sec
60	-	0.01	1000	ON-100	-3.017	09 min 58 sec
75	0.01	0.01	1000	OFF	-1.689	29 min 40 sec
75	-	0.01	1000	ON-100	-3.084	21 min 57 sec
120	0.01	0.01	1000	OFF	-1.738	31 min 16 sec
120	-	0.01	1000	ON-100	-3.309	20 min 13 sec

Note: The prior alpha governs whether the documents in LDA contain an exclusive (smaller values) or broad (larger values) range of topics. The prior beta determines whether the words in LDA belong to many (larger values) or only a few topics (smaller values). For the models with optimization “ON”, the prior alpha was optimized for each topic separately. Hence the absence of an overall alpha value for the respective models. The number of iterations determines how many iterations the LDA uses to find the optimal topic-document and word-topic distributions, whereby 1000 iterations are the default value. The optimization interval sets the length of the search interval of hyperparameter optimization. It was set to 100 for all models. Models without hyperparameter optimization have no corresponding value.

* The lowest coherence score is marked by an asterisk.