Finally, we investigated the benefit of latent topic meta-information. We ran xLDA estimating latent topic metadata ($x$LDAmtic) on the Wikipedia dataset with and without co-citation relations assuming 10 topics. We show the estimated covariance matrices and qualitatively compare the correlations found with the topics found.

**Results:** The perplexity results on Cora, Fig. 2(a), clearly show that xLDA can significantly be less uncertain about the remaining words than LDA and DMR ($K = 25$). The reason is that after seeing a few words in one topic, xLDA uses the link structure to infer that words in a related topic may also be probable. In contrast, LDA cannot predict the remaining words as well until a large portion of the document has been observed so that all of its topics are represented. Only when a very small number of words have been observed, the difference starts to vanish. This performance gain was also very stable when varying the number of
(a) Perplexity (the lower, the better) for different percentages of observed words per document.

(b) Perplexity using 70% / 30% training/test splits for different numbers of topics.

(c) Hellinger distance (the lower, the better) of linked documents for different number of topics.

(d) Average link log likelihood (the higher, the better) for different number of topics.

Fig. 3. Results on the Wikipedia: Perplexity, Hellinger distance, and average link log-likelihood for LDA, DMR id, DMR ra, and xLDA using co-citation. For the average link log-likelihood, we also compare to RTM. (Best viewed in color.)

topics as shown in Fig. 2(b). For larger numbers of topics, LDA starts to break down compared to DMR and xLDA. This effect can be broken when optimizing the Dirichlet hyperparameters for each document separately as essentially done by DMR id. Again, however, xLDA can make use of the link structure to infer that words in a related topic may also be probable. That xLDA captures the link structure better is best seen when considering the Hellinger distances between co-linked documents as shown in Fig. 2(c). Most surprisingly, however, in predicting links based on the topics proportions only, xLDA’s performance is even comparable with RTM’s, a recent fully-generative model.

The perplexity results on Wikipedia, Fig. 3(a), show a similar result. When a small number of words have been observed, there is less uncertainty about the remaining words under DMR and xLDA than under LDA (K = 25). Given that this dataset is much smaller, we can better observe that LDA cannot