

## A Datasets

For training our model, we use the WN18 knowledge graph that is derived from WordNet. It consists of 40,943 nodes and 151,442 edges (Bordes et al., 2014). We also use these edges as unlabeled edges. The following text datasets are considered in our experiments: (1) Corpus of Linguistic Acceptability (CoLA) (Warstadt et al., 2019) which contains 8,550 samples for training, 1,042 samples for validation, and 1,061 samples for testing; (2) Stanford Sentiment Treebank (SST-2) (Socher et al., 2013) that is a dataset of movie reviews consisting of 67,349 samples for training, 1,821 samples for validation, and 872 samples for testing; (3) The Brown University Standard Corpus of Present-Day American English (Brown) (Francis and Kucera, 1979) which includes American English texts printed in 1961, consisting of 15 different text categories. However, only 5 categories are used in this work. We split the corpus to 10,298 samples for training, 3,432 samples for validation, and 3,434 samples for testing.

## B Qualitative results

Table 4 includes some examples of MLM task performed by MG-BERT(base) and BERT, and demonstrates the importance of that real-world information of knowledge graph and global information of co-occurrence graphs in compensating BERT’s shortcomings.

<b>Input sentence</b>	<b>Expected</b>	<b>Top 5 predictions</b>	<b>Model</b>
fear dot [MASK] is so rambling and disconnected. it never builds any suspense.	com	com, #ter, #ting, #h, #han	MG-BERT
		#ing, #ting, #h, #ant, #ter	BERT
how do you spell clich [MASK]	?	?, ), (, ', !	MG-BERT
		t, d, s, r, '	BERT
it can not be enjoyed , even on the level that one enjoys a bad slasher flick , primarily because it is [MASK].	dull	bad, funny, awful, horrible, dull	MG-BERT
		bad, funny, enjoyable, entertaining, predictable	BERT
a giggle a [MASK]	minute	laugh, minute, giggle, chuckle, little	MG-BERT
		laugh, little, giggle, lot, bit	BERT
this film was made to get laughs from the slowest person in the [MASK], just pure slapstick with lots of inane, inoffensive screaming and exaggerated facial expressions.	audience	world, universe, room, audience, theater	MG-BERT
		world, universe, room, movie, film	BERT

Table 4: Some examples from SST-2 dataset.