

Table 1 Performance comparison: homogeneous KG versus heterogeneous KG (* means the performance difference is statistically significant, TC-Acc means the triplet classification accuracy).

Comparison Configuration	Top-5 Acc (Raw)	Top-5 Acc (Filtered)	Top-10 Acc (Raw)	Top-10 Acc (Filtered)	MRR (Raw)	MRR (Filtered)	MAP	TC-Acc
CWE-only KG	0.456	0.592	0.574	0.681	0.308	0.449	0.564	0.778
Full KG	0.473*	0.628*	0.596*	0.713*	0.316	0.459*	0.576*	0.792*
CAPEC-only KG	0.753	0.853	0.857	0.881	0.572	0.782	0.773	0.637
Full KG	0.777*	0.882*	0.893*	0.914*	0.589	0.793*	0.787*	0.662*
CWE&CAPEC KG	0.471	0.614	0.599	0.701	0.327	0.471	0.606	0.746
Full KG	0.487*	0.638*	0.612*	0.717*	0.337*	0.494*	0.625*	0.770*

Table 2 p-Value of Statistical Test: homogeneous KG versus heterogeneous KG.

	Top-5 Acc (Raw)	Top-5 Acc (Filtered)	Top-10 Acc (Raw)	Top-10 Acc (Filtered)	MRR (Raw)	MRR (Filtered)	MAP	TC-Acc
CWE-only KG	0.005859	0.003906	0.008004	0.001953	0.064453	0.011407	0.021303	0.001953
CAPEC-only KG	0.009091	0.005889	0.005889	0.005761	0.052773	0.040555	0.028954	0.005510
CWE&CAPEC KG	0.009766	0.013672	0.013672	0.003906	0.003906	0.001953	0.003906	0.001953

Table 3 Link prediction - overall performance (* means M-7 performs statistically significantly better than M-1).

Model	Top-5 Acc (Raw)	Top-5 Acc (Filtered)	Top-10 Acc (Raw)	Top-10 Acc (Filtered)	MRR (Raw)	MRR (Filtered)	MAP
M-1	0.412	0.503	0.507	0.583	0.292	0.402	0.591
M-2	0.392	0.474	0.496	0.558	0.282	0.391	0.573
M-3	0.418	0.528	0.519	0.596	0.311	0.405	0.606
M-4	0.420	0.542	0.533	0.612	0.314	0.420	0.604
M-5	0.433	0.529	0.533	0.612	0.317	0.419	0.624
M-6	0.447	0.573	0.542	0.640	0.325	0.448	0.647
M-7	0.465*	0.596*	0.560*	0.661*	0.340*	0.464*	0.688*

Table 4 p-Value of Statistical Test: M-7 versus M-1 on overall performance of link prediction.

Top-5 Acc (Raw)	Top-5 Acc (Filtered)	Top-10 Acc (Raw)	Top-10 Acc (Filtered)	MRR (Raw)	MRR (Filtered)	MAP
0.013794	0.001953	0.003906	0.001953	0.005761	0.005510	0.005761

Table 5 Link prediction - Top-5 Accuracy (filtered) by relationship type and predict-head/tail (* means M-7 performs statistically significantly better than M-1).

	Model	instanceof	targetof	childof	parentof	peerof	canfollow	canprecede	semantic	overall
Predict Head < ?h, r, t >	M-1	0.024	0.396	0.572	0.792	0.111	0.286	0.571	0.193	0.378
	M-2	0.017	0.386	0.555	0.758	0.056	0.286	0.571	0.189	0.359
	M-3	0.047	0.409	0.578	0.811	0.111	0.307	0.653	0.204	0.383
	M-4	0.093	0.420	0.618	0.845	0.167	0.357	0.653	0.234	0.399
	M-5	0.065	0.420	0.630	0.831	0.167	0.429	0.643	0.259	0.396
	M-6	0.141	0.443	0.684	0.858	0.278	0.457	0.669	0.294	0.441
	M-7	0.193*	0.466*	0.703*	0.865*	0.278*	0.500*	0.714*	0.314*	0.469*
Predict Tail < h, r, ?t >	M-1	0.650	0.342	0.789	0.568	0.222	0.357	0.591	0.437	0.625
	M-2	0.595	0.330	0.760	0.557	0.167	0.357	0.571	0.429	0.588
	M-3	0.822	0.342	0.808	0.575	0.222	0.371	0.623	0.442	0.671
	M-4	0.831	0.352	0.851	0.604	0.222	0.428	0.714	0.465	0.684
	M-5	0.729	0.355	0.845	0.656	0.278	0.428	0.714	0.477	0.662
	M-6	0.838	0.386	0.862	0.656	0.278	0.457	0.757	0.500	0.704
	M-7	0.848*	0.398*	0.878*	0.689*	0.278*	0.500*	0.766*	0.523*	0.720*

Table 6 p-value of Statistical Test: M-7 versus M-1 on overall performance Top-5 Accuracy (filtered) by relationship type and predict-head/tail of link prediction.

	instanceof	targetof	childof	parentof	peerof	canfollow	canprecede	semantic	overall
Head	0.001953	0.014369	0.009091	0.009030	0.034611	0.014147	0.034097	0.009091	0.005761
Tail	0.001953	0.009091	0.005666	0.001953	0.009091	0.014029	0.005857	0.005761	0.013672

Table 7 Triplet classification - detailed and overall performance (* means M-7 performs statistically significantly better than M-1).

Model	instanceof	targetof	childof	parentof	peerof	canfollow	canprecede	semantic	overall
M-1	0.837	0.795	0.682	0.740	0.694	0.714	0.679	0.866	0.781
M-2	0.834	0.756	0.676	0.736	0.639	0.652	0.607	0.866	0.770
M-3	0.842	0.812	0.703	0.747	0.722	0.750	0.714	0.877	0.792
M-4	0.873	0.813	0.713	0.760	0.722	0.750	0.714	0.877	0.809
M-5	0.857	0.818	0.727	0.764	0.694	0.750	0.714	0.889	0.807
M-6	0.884	0.831	0.730	0.777	0.722	0.786	0.714	0.889	0.818
M-7	0.911*	0.834*	0.747*	0.797*	0.756*	0.791*	0.730*	0.889*	0.829*

Table 8 p-value of Statistical Test: M-7 versus M-1 on detailed and overall performance of triplet classification.

instanceof	targetof	childof	parentof	peerof	canfollow	canprecede	semantic	overall
0.005793	0.013911	0.009091	0.005541	0.009091	0.005793	0.005761	0.007882	0.005793

Table 9 $\langle CVE, instanceof, CWE \rangle$ Prediction performance for newly published CVEs (* means M-7 performs statistically significantly better than M-1).

Model	Top-5 Acc	Top-10 Acc	MRR	MAP	TC-Acc
M-1	0.446	0.773	0.281	0.281	0.786
M-2	0.426	0.687	0.275	0.275	0.774
M-3	0.717	0.822	0.498	0.498	0.796
M-4	0.729	0.833	0.565	0.565	0.802
M-5	0.677	0.790	0.433	0.433	0.788
M-6	0.756	0.837	0.574	0.574	0.838
M-7	0.771*	0.857*	0.607*	0.607*	0.864*

Table 10 p-value of Statistical Test: M-7 versus M-1 on prediction performance of $\langle CVE, instanceof, CWE \rangle$ for newly published CVEs.

Top-5 Acc	Top-10 Acc	MRR	MAP	TC-Acc
0.001953	0.003906	0.003906	0.001953	0.005761