

	original	KPG
avg tree size	183.4	112.5
avg noise size	9.9	2.9
avg noisy ratio	5.90%	2.60%

Table 1: The analysis about noisy comments on Weibo22, where *original* denotes the input propagation graphs and *KPG* denotes the generated key graphs via our KPG.

	cascade size		max depth		max degree		number of nodes with degree>1	
	Non-rumor	Rumor	Non-rumor	Rumor	Non-rumor	Rumor	Non-rumor	Rumor
mean	311.74	149.19	2.89	3.22	249.93	66.1	16.83	20.72
min	1	1	1	1	0	0	0	0
median	5	9	2	3	3	6	1	1
max	30791	26535	17	29	22348	7805	1954	3705

Table 2: Statistics of Weibo22.



Figure 1: Topics of COVID-19 related rumors.



Figure 2: Publishers' verification reasons of COVID-19 related rumors.



Figure 3: Publishers' verification reasons of other rumors.

time delta (mins)	root	20	60	120	240	full
KPG	0.832	0.875	0.878	0.881	0.881	0.889
BERT	0.747	0.750	0.758	0.760	0.748	0.753

Table 3: The results of KPG and BERT on Twitter16 under different stages of spread, where *root* represents using only root post for the veracity, and *full* means using full propagation graphs.

		Acc.	NR F_1	FR F_1	TR F_1	UR F_1
Twitter15	GPT-3.5	0.313	0.084	0.188	0.410	0.368
	GPT-4	0.419	0.535	0.376	0.261	0.431
	KPG	0.893	0.921	0.898	0.903	0.847
Twitter16	GPT-3.5	0.375	0.109	0.172	0.520	0.434
	GPT-4	0.394	0.498	0.369	0.266	0.397
	KPG	0.889	0.894	0.836	0.930	0.896

Table 4: The results of LLMs on Twitter15 and Twitter16.

		Acc.	R F_1	NR F_1
PHEME	GPT-3.5	0.599	0.484	0.672
	GPT-4	0.615	0.494	0.689
	KPG	0.859	0.863	0.854
Weibo22	GPT-3.5	0.670	0.704	0.627
	GPT-4	0.745	0.780	0.696
	KPG	0.949	0.949	0.948

Table 5: The results of LLMs on PHEME and Weibo22.