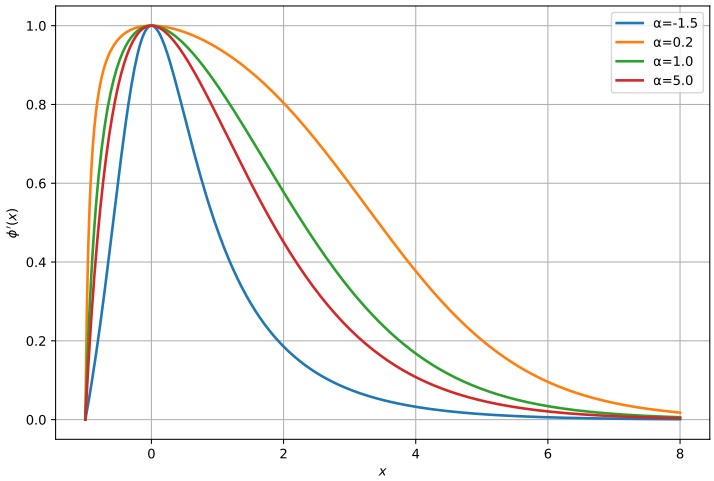

Figure 1: The smooth influence function $\phi'(x) = \frac{(1+\alpha)(x+1)}{(x+1)+\alpha\exp(x)}$, $x \geq -1$ under different parameter settings α.

Table 1: Performance comparison of Residual TPP based on two different influence functions (denoted as "old"/"new" respectively) on example benchmark datasets: MIMIC-II and Earthquake. For the new influence function, we set $\alpha = 1$. We evaluate the model's *goodness-of-fit* (Log-Likelihood, higher is better) and *prediction performance* (Time RMSE / Type Error Rate, lower is better).

Model	MIMIC-II		Earthquake	
	L-L	Time/Type	L-L	Time/Type
MHP	-2.839	0.925/27.9%	-4.155	1.475/60.6%
RMTPP	-2.626	0.998/37.8%	-4.643	1.956/52.9%
Res RMTPP(old)	**-2.045**	**0.915**/26.7%	-3.689	1.420/**52.7%**
Res RMTPP(new)	-2.150	0.923/**24.4%**	**-3.667**	**1.415**/52.8%
NHP	-2.031	1.010/26.7%	-2.389	1.910/53.9%
Res NHP(old)	**-1.825**	0.913/20.9%	-1.930	**1.416/52.8%**
Res NHP(new)	-2.040	**0.900/18.0%**	**-1.916**	1.419/52.9%
SAHP	-4.672	0.971/23.8%	-3.338	1.463/54.1%
Res SAHP(old)	**-4.488**	0.935/**18.0%**	-3.335	**1.455**/53.3%
Res SAHP(new)	-4.571	**0.925**/22.1%	**-3.312**	1.458/**53.0%**
THP	-2.048	1.129/35.5%	-3.498	1.857/54.7%
Res THP(old)	**-2.040**	0.930/27.9%	**-3.415**	**1.403/52.8%**
Res THP(new)	-2.879	**0.911/26.7%**	-3.533	1.417/52.8%
AttNHP	-2.500	1.030/35.4%	-2.896	1.822/54.5%
Res AttNHP(old)	**-2.197**	0.932/36.0%	-3.147	**1.413/52.8%**
Res AttNHP(new)	-2.918	**0.914/32.0%**	**-1.969**	1.419/52.8%
ODETPP	-1.855	1.416/22.1%	-2.203	2.396/56.0%
Res ODETPP(old)	**-1.371**	0.934/**19.2%**	-2.340	1.412/**53.0%**
Res ODETPP(new)	-2.365	**0.921**/19.3%	**-1.889**	**1.411**/53.1%

Table 2: End-to-end training time for neural TPPs and their corresponding Residual TPPs. The first row, "MHP + RED", reports the combined runtime for Steps 1 and 2, representing the mean computational time across 10 independent trials. For each Residual TPP, the training parameters are kept the same as in the baseline model, and the result reflects the total time cost of the 3-step procedure (Hawkes fitting + RED filtering + and neural TPP training). All training was conducted on a CPU.

Model	End-to-end Runtime (seconds)					
	MIMIC-II	Retweet	Earthquake	StackOverflow	Amazon	Volcano
MHP+RED	1.42	0.27	0.11	0.84	1.49	0.02
RMTPP	**9.70**	195.3	26.00	71.20	95.75	34.20
Res RMTPP	9.72	**188.7**	**20.01**	**70.44**	**95.19**	**26.42**
NHP	34.70	492.6	25.32	206.3	256.6	77.05
Res NHP	**29.52**	**441.9**	**20.57**	**190.8**	**245.1**	**54.47**
SAHP	128.5	498.4	24.90	498.0	505.8	37.30
Res SAHP	**113.4**	**434.9**	**16.96**	**461.9**	**450.3**	**24.42**
THP	**10.45**	1183	19.20	81.20	257.7	39.90
Res THP	10.67	**1029**	**17.77**	**76.44**	**244.8**	**29.82**
AttNHP	68.60	9475	162.2	1924	6375	1093
Res AttNHP	**52.92**	**7195**	**143.1**	**1863**	**5646**	**657.5**
ODETPP	9.96	106.4	42.00	243.6	196.4	51.96
Res ODETPP	**9.64**	**102.7**	**37.43**	**224.2**	**192.1**	**40.88**

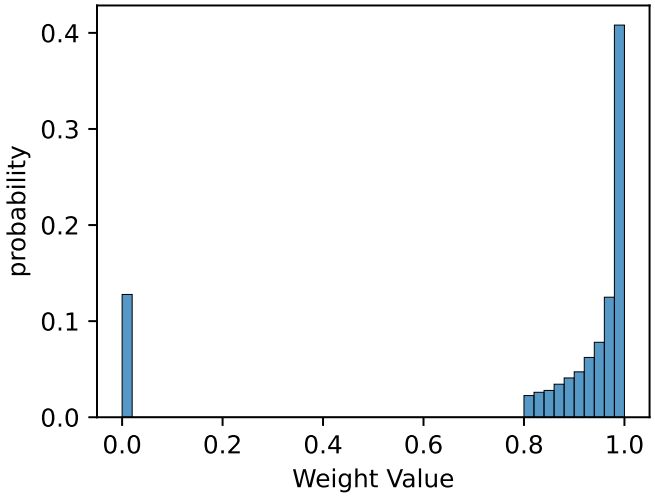

Figure 2: Distribution of weight values on the simulated Hawkes dataset. The dataset consists of 300 sequences generated by a 1D Hawkes process with the intensity function $\lambda(t) = 0.2 + 0.6 \int_0^t e^{-1.2s} dN(s)$. Each sequence contains an average of 36 events, with 200 sequences used for training and 100 for testing. The RED technique was applied with the same parameter settings as in the paper. The histogram displays the distribution of weight values calculated by RED for all events across all sequences in the dataset.