Supplementary Material: Learning Multivariate Hawkes Process via Graph Recurrent Neural Network

Anonymous Author(s)

A DATASET

A.1 Graph Construction

We divide our real-world dataset into two categories as geometric and web data. The geometric dataset have an event stream that can be influenced by nearby regions. For example, the next region that a taxi picks up is dependent on the last event and the accessibility of regions. *NYC TAXI*, *Earthquake*, *911 Calls* dataset belongs to geometric data. The web dataset has an event stream in the form of posting with posting time and the topics of posting. It includes Reddit and StackOverflow.

Table 1: The data statistics of GRTPP for 5 dataset.

Dataset	Event Instances	E types (nodes)	Edges	Avg. Length
NYC TAXI	100,000	299	7474	38
Reddit	192,864	100	9694	104
Stack Overflow	480,413	22	231	72
Earthquake	256,932	69	8100	500
911 Calls	290,293	82	272	403

A.1.1 Geometric Data.

- *NYC TAXI* Taxi demand patterns vary depending on the region of the city. We construct a graph to effectively represent such spatial relationships between cities. We define each city as a node in the graph. For all cities, we add edges between cities that are the most twenty-five closest. We used the longitude and latitude information to compute the distance between cities.
- *Earthquake* Earthquake can sporadically occur. When it first occurs, the aftershock will be derived based on the distance from the origin. We assign earthquake occurrence to the closest observatory (event nodes). We add edges between the city that is the ten closest observatories. We used the longitude and latitude information to compute the distance between observatories.
- 911 Calls For 911 calls, we assume that a accident in certain cause several calls from the regions from the place. We defines all calls from certain regions as event and assign it based on zip code that the city is contained. We add edges between city that is the most eight closest cities as NYC TAXI construct graph.

A.1.2 Web Data.

- *Reddit* The positing in Reddit share several user interests. We assume that a user who likes soccer is likely to be interested in baseball. Thus, we construct the relational information graph based on the topic similarity of subreddits. We define all subreddits as event nodes. Then, we compute node similarity based on raw features and use it to construct a k-nearest neighbors graph. Thus, the relation graph of Reddit includes the event occurrence in similar subreddits.
- StackOverflow For Stackoverflow, the definition of an event is the type of obtained badge when a user answers some questions. Since it is difficult to define a graph in such a situation, Thus, we use a fully-connected graph.

B HYPERPARAMETERS

Table 2: The hyperparameter of GRTPP for 5 dataset.

Dataset	NYC TAXI	Reddit	Stack Overflow	Earthquake	911 Calls
hidden dimension	32	64	16	64	32
hidden layers	2	2	2	2	2
batch size	16	32	8	16	16

To guarantee the reproducibility of our works, we run our implementation several times to fill in the main performance table. For five datasets, we did a grid search for hyperparameter tuning, and hyperparameters for each dataset are described in table 3. We select the number of stacking GNN layers to be 2. It shows that the optimal estimation for event likelihood is achieved by the 2-hop neighbor event. For the number of negatives, we set it to be 8 for all dataset