Table 3: TFT configuration

|  |  |
| --- | --- |
| **Details** | **Values Chosen** |
| Dataset |
| Number of Entities generated | 400 |
| Number of samples selected | 600k |
| Parameters for TFT Network  |
| Size of State | 160 |
| Number of attention heads | 4 |
| Drop Rate | 0.1 |
| Number of hidden layers | 8 |
| Number of consecutive hidden layers | 4 |
| Parameters for Training  |
| Learning rates | 0.001,0.0001,0.01 |
| Size of Minibatch | 64 |
| Max Gradient Norm | 0.01 |