Discovering Frequency Bursting Patterns in Temporal Graphs – Parameters Settings

τ is an important parameter in our model and controls whether two interactions are related. To learn the appropriate τ, we pick a sample of 1000 nodes proportional to their frequencies of interactions. This is necessary since a large portion of the users are dormant. For each selected node, we extract the subgraph of radius τ around it. We can obtain the average growth rate in the subgraph sizes as τ is varied in a range $[t_{min}, t_{max}]$, where

$$Coverage(\tau) = \frac{subgraph\ \ size\ \ at\ \ \tau}{subgraph\ \ size\ \ at\ \ t_{max}} \tag{1}$$

In $Enron$, no clear pattern is visible as the growth rate is linear. We vary τ from 1 to 4 as that in [18]. In $Panama$, the coverage shows two jumps at $\tau = 120$ and $\tau = 480$. Thus, a threshold between 120 to 480 is a reasonable value for τ. In $Citation$, the coverage shows two jumps at $\tau = 1$ and $\tau = 4$. Thus, a threshold between 1 to 4 is a reasonable value for τ. In $Moveliens$, the growth rate is also linear, we vary τ from 12 to 48. Interactions in $Bitcoin$ remain active for a much longer while since a deal will go on for a long time. Thus, we vary τ from 3 to 12. Note that for each dataset, we use the second value of the parameter as default.

θ is set by domain scientists based on domain knowledge. In $Enron$ and $Citation$, we use the same setting as that in [8]. In $Panama$, we vary θ from 180 to 720 with a default 360 since it needs at least half year to finish a financial activity for the companies in Panama. In $Moveliens$, we vary θ from 30 to 120 with a default 60. This is because user activity will be significantly enhanced in the online movie recommendation system after a month of the release of the new movie. In $Bitcoin$, we vary θ from 7 to 28 with a default 14.