

A SUPPLEMENTARY MATERIAL

A.1 MAP ESTIMATION DETAILS

As described in Section 4.3, we perform maximum a posteriori (MAP) inference to estimate the parameters in all the discussed models. In this section, we present the MAP estimation details for the HP and DLS models by deriving the closed form expressions of the log-posterior function and its gradients; the optimization can then be carried out using L-BFGS-B (Byrd et al., 1995). The derivations for the PLS, BLS, RLS models follow analogously, since they can all be viewed as degenerate cases of the DLS model.

Before presenting the MAP estimation details, recall that the observed data $\{(u, v, \mathcal{H}_{uv})\}_{u,v \in V}$ are collected over a time period $[0, T]$, where $\mathcal{H}_{uv} \triangleq \{t_i^{uv}\}_{i=1}^{n_{uv}}$ records the set of all time-points at which u sent v a message.

A.1.1 Hawkes Process (HP) Model

Recall the Hawkes Process (HP) model:

$$\begin{aligned} \lambda_{uv}(t) &= \gamma + \sum_{k: t_k^{vu} < t} \sum_{b=1}^B \xi_b \phi_b(t - t_k^{vu}) & \forall u \neq v \\ N_{uv}(\cdot) &\sim \text{HawkesProcess}(\lambda_{uv}(\cdot)) & \forall u \neq v \end{aligned}$$

Notice that

$$\Lambda_{uv}(0, T) = \int_0^T \lambda_{uv}(t) dt = \gamma T + \sum_{b=1}^B \xi_b \sum_{k=1}^{n_{vu}} [\Phi_b(T - t_k^{vu}) - \Phi_b(0)]$$

where $\Phi_b(t) \triangleq \int_0^t \phi_b(s) ds$.

Placing Gamma(1, 1) priors on γ and each ξ_b , and denoting $\boldsymbol{\xi} \triangleq \{\xi_b\}_{b=1}^B$, the joint density can be written as

$$p(\{\mathcal{H}_{uv}\}_{u,v=1}^n, \gamma, \boldsymbol{\xi}) \propto \prod_{\substack{u,v=1 \\ u \neq v}}^n \left\{ e^{-\Lambda_{uv}(0, T)} \prod_{k=1}^{n_{uv}} \lambda_{uv}(t_i^{uv}) \cdot e^{-\gamma} \cdot \prod_{b=1}^B e^{-\xi_b} \right\}$$

and the log-posterior function is given by

$$\begin{aligned} \log p(\gamma, \boldsymbol{\xi} | \{\mathcal{H}_{uv}\}_{u,v=1}^n) &= \sum_{\substack{u,v=1 \\ u \neq v}}^n \left\{ -\Lambda_{uv}(0, T) + \sum_{i=1}^{n_{uv}} \log \lambda_{uv}(t_i^{uv}) \right\} - \gamma - \sum_{b=1}^B \xi_b \\ &= \sum_{\substack{u,v=1 \\ u \neq v}}^n \left\{ -\gamma T - \sum_{b=1}^B \xi_b \Delta_{b,T}^{vu} + \sum_{i=1}^{n_{uv}} \log \left(\gamma + \sum_{b=1}^B \xi_b \delta_{b,i}^{uv} \right) \right\} - \gamma - \sum_{b=1}^B \xi_b \end{aligned}$$

where $\langle \cdot, \cdot \rangle$ denotes the Euclidean inner-product, and we have adopted the shorthand notations

$$\begin{aligned} \Delta_{b,T}^{vu} &\triangleq \sum_{k=1}^{n_{vu}} [\Phi_b(T - t_k^{vu}) - \Phi_b(0)] \\ \delta_{b,i}^{uv} &\triangleq \sum_{k: t_k^{vu} < t_i^{uv}} \phi_b(t_i^{uv} - t_k^{vu}) \end{aligned}$$

to denote data statistics that can be pre-computed and cached for each pair of nodes $u, v \in V$ and kernel ϕ_b .

The gradients of the log-posterior are given by

$$\begin{aligned}\frac{\partial \log p}{\partial \gamma} &= -(n^2 - n)T + \sum_{\substack{u,v=1 \\ u \neq v}}^n \sum_{i=1}^{n_{uv}} \left(\gamma + \sum_{b=1}^B \xi_b \delta_{b,i}^{uv} \right)^{-1} - 1 \\ \frac{\partial \log p}{\partial \xi_b} &= \sum_{\substack{u,v=1 \\ u \neq v}}^n \left[-\Delta_{b,T}^{vu} + \sum_{i=1}^{n_{uv}} \delta_{b,i}^{uv} \left(\gamma + \sum_{b=1}^B \xi_b \delta_{b,i}^{uv} \right)^{-1} \right] - 1.\end{aligned}$$

A.1.2 Hawkes Dual Latent Space (DLS) Model

Recall the Hawkes Dual Latent Space (DLS) model:

$$\begin{aligned}\mathbf{z}_v &\sim \mathcal{N}(\mathbf{0}, \sigma^2 \mathbf{I}_{d \times d}) && \forall v \in V \\ \boldsymbol{\mu}_v &\sim \mathcal{N}(\mathbf{0}, \sigma_\mu^2 \mathbf{I}_{d \times d}) && \forall v \in V \\ \boldsymbol{\varepsilon}_v^{(b)} &\sim \mathcal{N}(\mathbf{0}, \sigma_\varepsilon^2 \mathbf{I}_{d \times d}) && \forall v \in V, b = 1, \dots, B \\ \mathbf{x}_v^{(b)} &\sim \boldsymbol{\mu}_v + \boldsymbol{\varepsilon}_v^{(b)} && \forall v \in V, b = 1, \dots, B \\ \lambda_{uv}(t) &= \gamma e^{-\|\mathbf{z}_u - \mathbf{z}_v\|_2^2} + \sum_{k: t_k^{vu} < t} \sum_{b=1}^B \beta e^{-\|\mathbf{x}_u^{(b)} - \mathbf{x}_v^{(b)}\|_2^2} \phi_b(t - t_k^{vu}) \\ N_{uv}(\cdot) &\sim \text{HawkesProcess}(\lambda_{uv}(\cdot)) && \forall u \neq v\end{aligned}$$

Placing Gamma(1, 1) priors on γ and β , setting $\sigma^2 = \sigma_\mu^2 = \sigma_\varepsilon^2 = 1$, and integrating out $\{\boldsymbol{\mu}_v\}_{v=1}^n$, the log-density function can be written as

$$\begin{aligned}\log p(\gamma, \beta, \{\mathbf{z}_v\}_{v=1}^n, \{\{\mathbf{x}_v^{(b)}\}_{b=1}^B\}_{v=1}^n \mid \{\mathcal{H}_{uv}\}_{u,v=1}^n) \\ = \sum_{\substack{u,v=1 \\ u \neq v}}^n \left\{ -\gamma e^{-\|\mathbf{z}_u - \mathbf{z}_v\|_2^2} T - \beta \sum_{b=1}^B \Delta_{b,T}^{vu} e^{-\|\mathbf{x}_u^{(b)} - \mathbf{x}_v^{(b)}\|_2^2} + \sum_{i=1}^{n_{uv}} \log \left(\gamma e^{-\|\mathbf{z}_u - \mathbf{z}_v\|_2^2} + \beta \sum_{b=1}^B \delta_{b,i}^{uv} e^{-\|\mathbf{x}_u^{(b)} - \mathbf{x}_v^{(b)}\|_2^2} \right) \right\} \\ - \frac{1}{2} \sum_{v=1}^n \sum_{b=1}^B \|\mathbf{x}_v^{(b)}\|_2^2 + \frac{B^2}{2(B+1)} \sum_{v=1}^n \|\bar{\mathbf{x}}_v\|_2^2 - \frac{1}{2} \sum_{v=1}^n \|\mathbf{z}_v\|_2^2 - \gamma - \beta\end{aligned}$$

where $\bar{\mathbf{x}}_v \triangleq \frac{1}{B} \sum_{b=1}^B \mathbf{x}_v^{(b)}$ denotes the mean latent position of node v across all basis-kernels.

The gradients of the log-posterior are given by

$$\begin{aligned}\frac{\partial \log p}{\partial \gamma} &= \sum_{\substack{u,v=1 \\ u \neq v}}^n \left[-T e^{-\|\mathbf{z}_u - \mathbf{z}_v\|_2^2} + \sum_{i=1}^{n_{uv}} e^{-\|\mathbf{z}_u - \mathbf{z}_v\|_2^2} h^{-1}(u, v, i) \right] - 1 \\ \frac{\partial \log p}{\partial \beta} &= \sum_{\substack{u,v=1 \\ u \neq v}}^n \sum_{b=1}^B r(u, v, b) e^{-\|\mathbf{x}_u^{(b)} - \mathbf{x}_v^{(b)}\|_2^2} - 1 \\ \nabla_{\mathbf{z}_v} \log p &= \sum_{\substack{u=1 \\ u \neq v}}^n \left\{ \gamma \left[-2T + \sum_{i=1}^{n_{uv}} (h^{-1}(u, v, i) + h^{-1}(v, u, i)) \right] e^{-\|\mathbf{z}_u - \mathbf{z}_v\|_2^2} \cdot 2(\mathbf{z}_u - \mathbf{z}_v) \right\} - \mathbf{z}_v \\ \nabla_{\mathbf{x}_v^{(b)}} \log p &= \sum_{\substack{u=1 \\ u \neq v}}^n \left\{ \beta [r(u, v, b) + r(v, u, b)] e^{-\|\mathbf{x}_u^{(b)} - \mathbf{x}_v^{(b)}\|_2^2} \cdot 2(\mathbf{x}_v^{(b)} - \mathbf{x}_u^{(b)}) \right\} - \mathbf{x}_v^{(b)} + \frac{B}{B+1} \cdot \bar{\mathbf{x}}_v\end{aligned}$$

where

$$h(u, v, i) \triangleq \gamma e^{-\|\mathbf{z}_u - \mathbf{z}_v\|_2^2} + \beta \sum_{b=1}^B \delta_{b,i}^{uv} e^{-\|\mathbf{x}_u^{(b)} - \mathbf{x}_v^{(b)}\|_2^2}$$

$$r(u, v, b) \triangleq -\Delta_{b,T}^{vu} + \sum_{i=1}^{n_{uv}} \delta_{b,i}^{uv} h^{-1}(u, v, i).$$

A.2 ADDITIONAL EXPERIMENT RESULTS

A.2.1 Further Experiment on Static Link Prediction

In Section 5.3, we noted that the experiment setup for the static link prediction task did not yield standard errors for the AUC scores reported in Table 3, since there was only one training/test split. To investigate the statistical significance of the results, we conducted a follow-up experiment.

For each dataset, we computed confidence intervals by performing six trials on subsets of the data. Specifically, in the i -th trial, we let the training set to contain all events during the period between the $\lceil \frac{i-1}{10} \rceil$ -th and the $\lfloor \frac{i+2}{10} \rfloor$ -th event, and the test set to contain all events during the period between the $\lceil \frac{i+2}{10} \rceil$ -th and $\lfloor \frac{i+4}{10} \rfloor$ -th event. In this way, each trial used 30% training data and 20% test data, with the training and test data being non-overlapping.⁹ As in Section 5.3, we fitted the model on the training set, and performed link prediction on the test set. The results are shown in Table 4.¹⁰

Table 4: Static link prediction AUC scores and standard deviations.

Model	ENRON	EMAIL	FACEBOOK
PLS	0.510 (0.009)	0.496 (0.015)	0.491 (0.013)
BLS	0.510 (0.009)	0.496 (0.015)	0.491 (0.013)
RLS	0.439 (0.073)	0.386 (0.081)	0.456 (0.055)
DLS	0.864 (0.016)	0.934 (0.016)	0.892 (0.040)
Spectral	0.516 (0.020)	0.526 (0.032)	0.492 (0.021)
node2vec	0.749 (0.050)	0.953 (0.007)	0.935 (0.033)

By conducting two-sided t -tests at the 95% confidence level, we conclude that while DLS significantly outperforms node2vec on ENRON, their performance differences on EMAIL and FACEBOOK are not significant.

A.2.2 Visualization of the Inferred Node-Similarity Matrices

We visualize the estimated homophily and reciprocal latent spaces of the DLS model by computing the pair-wise similarities $e^{-\|\mathbf{z}_u - \mathbf{z}_v\|_2^2}$ for every pair of nodes $u, v \in V$, and then plotting a heat-map of the inferred similarity matrices. Figures 4, 5, and 6 show the heat-maps (colors on log-scale) for both the homophily latent space and the reciprocal latent spaces corresponding to the hourly (ϕ_1), daily (ϕ_2), weekly (ϕ_3) exponential kernels and the weekly locally periodic kernel (ϕ_4) on all three datasets. For each similarity matrix, we performed hierarchical clustering on the rows to obtain a node-ordering and accordingly permuted the rows and columns of the matrix simultaneously. Notice that the similarity matrices exhibit different clustering block-structures, indicating that the user-interaction patterns are quite different across the homophily and reciprocal latent spaces with different kernels and time-scales.

⁹Notice, however, that the training/test data across different trials may share common observations. Thus, strictly speaking, the trials are not independent, and the computed standard error estimates might under-estimate the "true" associated uncertainty.

¹⁰Note that the overall performance for all methods are slightly degraded since we are only using subsets of the data.

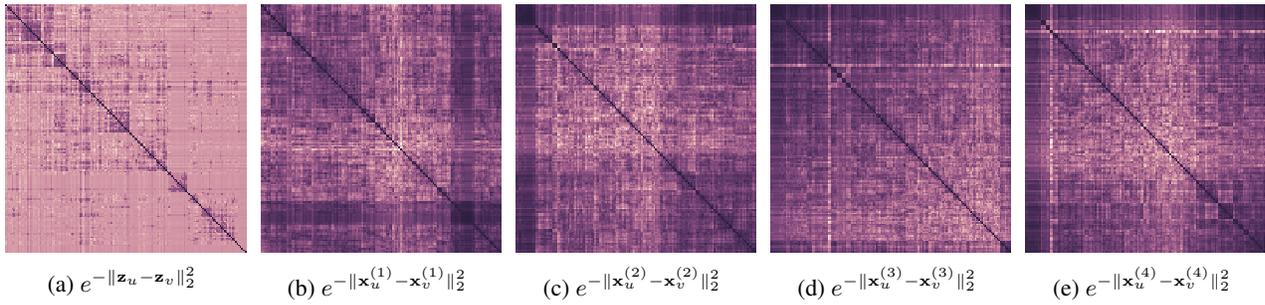


Figure 4: Inferred node-similarity matrices in ENRON.

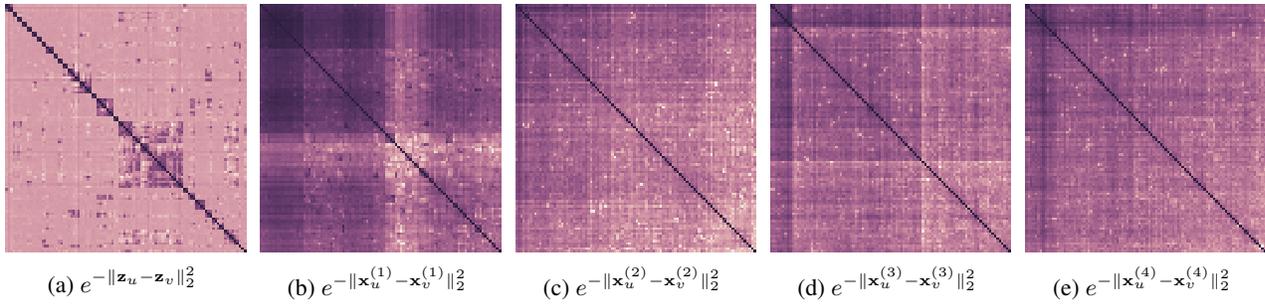


Figure 5: Inferred node-similarity matrices in EMAIL.

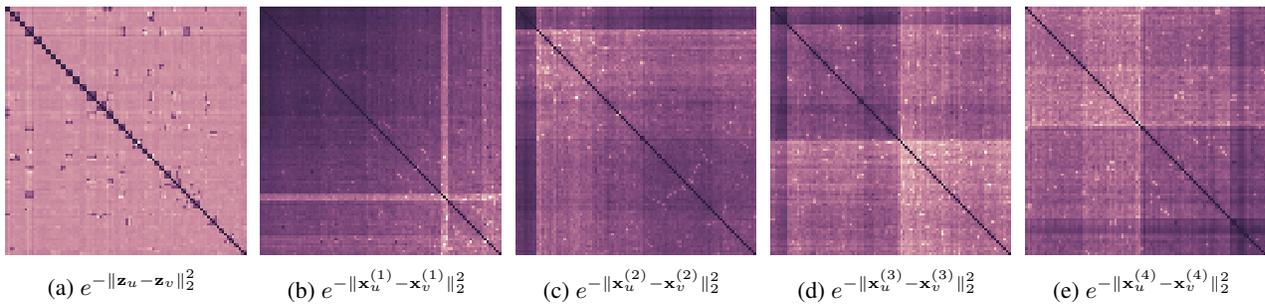


Figure 6: Inferred node-similarity matrices in FACEBOOK.