# Supplement for The Hawkes Edge Partition Model

## A INFERENCE

Next we shall explain the Gibbs sampling algorithm to infer the parameters of the Hawkes-EPM.

### A.1 GIBBS SAMPLING

The conditional intensity function of the Hawkes-EPM, for the directed events from $u$ to $v$, is

$$\begin{aligned}\lambda_{u,v}(t) &= \mu_{u,v} + \sum_{j:t_j\in\mathcal{H}_{v,k',k,u}(t)}\gamma_{k,k'}(t-t_j)\\ &= \sum_{k,k'}\left\{\mu_{u,k,k',v} + \sum_{j:t_j\in\mathcal{H}_{v,k',k,u}(t)}\alpha_{k,k'}\exp[-(t-t_j)/\delta]\right\}.\end{aligned} \tag{1}$$

**Sampling latent variables $\{z_i^s, z_i^d\}_{i=1}^N$:** For each event $(t_i, s_i, d_i)$, we utilize an auxiliary binary variable $b_i$ to denote whether $i\text{-th}$ event is triggered by the base rate (exogenous) or by opposite past interactions (endogenous) as

$$(b_i \mid -) \sim \text{Bernoulli}(\mu_{s_i,d_i}/\lambda_{s_i,d_i}(t_i)). \tag{2}$$

Then, we sample the latent patterns $(z_i^s, z_i^d)$ for each event as

$$(z_i^s, z_i^d \mid -) \sim \begin{cases}\text{Cat}\left(\frac{\{\mu_{s_i,k,k',d_i}\}_{k,k'=1}^K}{\lambda_{s_i,d_i}(t_i)}\right), & \text{if } b_i = 1\\ \text{Cat}\left(\frac{\{\breve{\lambda}_{s_i,k,k',d_i}(t_i)\}_{k,k'=1}^K}{\lambda_{s_i,d_i}(t_i)}\right), & \text{otherwise}\end{cases} \tag{3}$$

where $\text{Cat}(\cdot)$ denotes the categorical distribution, and we define

$$\breve{\lambda}_{s_i,k,k',d_i}(t_i) \equiv \sum_{j:t_j\in\mathcal{H}_{d_i,k',k,s_i}(t)}\alpha_{kk'}\exp[-\delta(t_i-t_j)]. \tag{4}$$

Given the sampled latent variables, we update the sufficient statistics as

$$\begin{aligned}\widehat{m}_{u,k,k',v} &\equiv \sum_j \mathbf{1}(b_j = 1, s_j = u, d_j = v, z_j^s = k, z_j^d = k'),\\ \breve{m}_{u,k,k',v} &\equiv \sum_j \mathbf{1}(b_j = 0, s_j = u, d_j = v, z_j^s = k, z_j^d = k').\end{aligned} \tag{5}$$

The log-posterior of the observed temporal events $\mathcal{D} \equiv \{(t_i, s_i, d_i)\}_{i=1}^N$ is shown in Eq. 6

$$\begin{aligned}\mathcal{L}(\Theta) =& \sum\nolimits_i \log\left\{\mu_{s_i,d_i} + \sum\nolimits_{k,k'}\sum\nolimits_{j:t_j\in\mathcal{H}_{d_i,k',k,s_i}(t_i)}\alpha_{kk'}\exp\left[-(t_i-t_j)/\delta\right]\right\}\\ &- \sum\nolimits_i\left\{\mu_{s_i,d_i}T + \sum\nolimits_{k,k'}\sum\nolimits_{j:t_j\in\mathcal{H}_{d_i,k',k,s_i}(t_i)}\alpha_{kk'}\delta(1-\exp\left[-(t_i-t_j)/\delta\right])\right\}\\ &+ \log\Pr(\Theta).\end{aligned} \tag{6}$$

**Sampling the kernel parameters $\{\alpha_{kk'}\}$:** As we place gamma priors over $\alpha_{kk'}$ as $\alpha_{kk'} \sim \text{Gamma}(1,1)$, and thus we have

$$\begin{aligned}(\alpha_{kk'} \mid -) \sim& \text{Gamma}\,(1 + \breve{m}_{\cdot,k,k',\cdot},\\ &1/\left[1 + \sum\nolimits_i \sum_{j:t_j\in\mathcal{H}_{d_i,k',k,s_i}(t_i)}\frac{1}{\delta}\left(1-\exp\left[-\frac{(T-t_j)}{\delta}\right]\right)\right]\Bigg),\end{aligned} \tag{7}$$

where $\check{m}_{\cdot k,k'\cdot} \equiv \sum_i \check{m}_{s_i,k,k',d_i}$, and $\check{m}_{\cdot k,k'\cdot}$ denotes the total number of endogenous events associated with the latent pattern $(k, k')$.

**Sampling the base intensity $\{\mu_{u,k,k',v}\}$:** As we have gamma prior over $\mu_{u,k,k',v}$ as $\mu_{u,k,k',v} \sim \mathrm{Gamma}(\tilde{\mu}_{u,k,k',v}, 1/(\exp[-\mathbf{x}_{u,v}^{\mathrm{T}}\boldsymbol{\beta}_{kk'}]))$, and thus we have

$$(\mu_{u,k,k',v} \mid -) \sim \mathrm{Gamma}\left(\tilde{\mu}_{u,k,k',v} + \widehat{m}_{u,k,k',v}, 1/(T + \exp[-\mathbf{x}_{u,v}^{\mathrm{T}}\boldsymbol{\beta}_{kk'}])\right), \tag{8}$$

Marginalizing out $\mu_{u,k,k',v}$ from the likelihood leads to

$$\begin{aligned}\Pr(\mathcal{D} \mid \mathbf{x}_{u,v}, \boldsymbol{\beta}_{kk'}) &= \int \Pr(\mathcal{D} \mid \mu_{u,k,k',v})\Pr(\mu_{u,k,k',v} \mid \mathbf{x}_{u,v}, \boldsymbol{\beta}_{kk'})\mathrm{d}\mu_{u,k,k',v} \\ &\propto \mathrm{NB}(\widehat{m}_{u,k,k',v}; \tilde{\mu}_{u,k,k',v}, \sigma[\mathbf{x}_{u,v}^{\mathrm{T}}\boldsymbol{\beta}_{kk'} + \log(T)]),\end{aligned}$$

where $\sigma(x) = 1/(1 + \exp(-x))$ denotes the logistic function, and $\mathrm{NB}(\cdot)$ denotes the Negative-Binomial distribution. Using the Pólya-Gamma data augmentation strategy (Zhou et al., 2012; Polson et al., 2013), we first sample

$$\begin{aligned}&(\omega_{u,k,k',v} \mid -) \sim \mathrm{PG}(\mu_{u,k,k',v} + \widehat{m}_{u,k,k',v}, \psi_{u,k,k',v}), \\ &(\psi_{u,k,k',v} \mid -) \sim \mathcal{N}(\mu_{\psi}, \sigma_{\psi}),\end{aligned} \tag{9}$$

where PG denotes a Pólya-Gamma draw, and where

$$\begin{aligned}\psi_{u,k,k',v} &\equiv \mathbf{x}_{uv}^{\mathrm{T}}\boldsymbol{\beta}_{kk'} + \log(T\pi_{uv}), \\ \pi_{uv} &\sim \log\mathcal{N}(0, \tau^{-1}) \\ \sigma_{\psi} &= [\omega_{u,k,k',v} + \tau]^{-1}, \\ \mu_{\psi} &= \sigma_{\psi}\left[(\widehat{m}_{u,k,k',v} - \mu_{u,k,k',v})/2 + \tau(\mathbf{x}_{uv}^{\mathrm{T}}\boldsymbol{\beta}_{kk'} + \log(T))\right],\end{aligned}$$

where $\log\mathcal{N}(\cdot)$ denotes the lognormal distribution.

**Sampling the regression coefficients $\{\boldsymbol{\beta}_{kk'}\}$:** Given $\{\boldsymbol{\psi}_{kk'} \equiv (\psi_{1kk'1}, \ldots, \psi_{Ukk'V})\}$, we sample $\{\boldsymbol{\beta}_{kk'}\}$ as

$$(\boldsymbol{\beta}_{k,k'} \mid -) \sim \mathcal{N}(\boldsymbol{\mu}_{\beta}, \boldsymbol{\Sigma}_{\beta}), \tag{10}$$

where $\boldsymbol{\Sigma}_{\beta} = (\tau\mathbf{X}^{\mathrm{T}}\mathbf{X} + \mathbf{A})^{-1}$, $\mathbf{A} \equiv \mathrm{diag}[\nu_1^{-1}, \ldots, \nu_D^{-1}]$, $\boldsymbol{\mu}_{\beta} = \tau\boldsymbol{\Sigma}_{\beta}\mathbf{X}^{\mathrm{T}}(\boldsymbol{\psi}_{kk'} - \log(T))$, and $\mathbf{X} \equiv [\mathbf{x}_{11}, \ldots, \mathbf{x}_{UV}]^{\mathrm{T}}$.

The full procedure of our Gibbs sampler is summarized in Algorithm 1.

**Algorithm 1** Gibbs Sampler for the Hawkes Edge Partition Model

```
Input: events data D = {(t_i, s_i, d_i)}_{i=1}^N, {Φ, Ω} inferred by the HGaP-EPM, maximum iterations J
Output: {μ_{u,k,k',v}}, {α_{kk'}}, {(z_i^s, z_i^d)}
for l = 1:J do
  for n = 1:N do
    Sample b_i (Eq. 2)
    Sample the latent variables (z_i^s, z_i^d) (Eq. 3)
    Update the intensity function λ_{u,v}(t_i) (Eq. 1)
  end for
  Update m̂_{u,k,k',v} and m̌_{u,k,k',v} (Eq. 5)
  Sample the base intensities {μ_{u,k,k',v}} (Eq. 8)
  Sample the parameters {β_{kk'}}, {ω_{u,k,k',v}}, {ψ_{u,k,k',v}} (Eqs. 10; 9)
  Sample the kernel parameters {α_{k,k'}} (Eq. 7)
end for
```

## B BASELINE MODELS

**The Hawkes Edge Partition Model (Hawkes-EPM)** For each pair of nodes $(u, v)$, $u, v \in \mathcal{V}$, and $u \neq v$,

$$\begin{aligned}
\mu_{u,k,k',v} &\sim \mathrm{Gamma}(\tilde{\mu}_{u,k,k',v}, 1/(\exp[-\mathbf{x}_{u,v}^{\mathrm{T}}\boldsymbol{\beta}_{kk'}])),\\
\tilde{\mu}_{u,k,k',v} &\equiv \phi_{u,k}\Omega_{k,k'}\phi_{v,k'},\\
\boldsymbol{\beta}_{k,k'} &\sim \mathcal{N}(\mathbf{0}, \mathbf{A}),\\
\alpha_{kk'} &\sim \mathrm{Gamma}(e_0, 1/f_0),\\
\lambda_{u,v}(t) &= \sum_{k,k'}\left\{\mu_{u,k,k',v} + \sum_{j:t_j\in\mathcal{H}_{v,k',k,u}(t)} \alpha_{kk'}\exp[-(t-t_j)/\delta]\right\},\\
N_{uv}(t) &\sim \text{Hawkes Process}(\lambda_{uv}(t)),
\end{aligned}$$

where $\mathbf{A} \equiv \mathrm{diag}[\nu_1^{-1}, \ldots, \nu_D^{-1}]$.

**The Hawkes Dual Latent Space (DLS) (Yang et al., 2017)** For each pair of nodes $(u, v)$, $u, v \in \mathcal{V}$, and $u \neq v$,

$$\begin{aligned}
\mathbf{z}_v &\sim \mathcal{N}(\mathbf{0}, \sigma^2 \mathrm{I}_{d\times d}),\\
\boldsymbol{\mu}_v &\sim \mathcal{N}(\mathbf{0}, \sigma_\mu^2 \mathrm{I}_{d\times d}),\\
\boldsymbol{\epsilon}_v^{(b)} &\sim \mathcal{N}(\mathbf{0}, \sigma_\epsilon^2 \mathrm{I}_{d\times d}),\\
\mathbf{x}_v^{(b)} &\sim \boldsymbol{\mu}_v + \epsilon_v^{(b)},\\
\lambda_{uv}(t) &= \phi\, e^{-\|\mathbf{z}_u-\mathbf{z}_v\|_2^2} + \sum_{j:t_j\in\mathcal{H}_{v,u}(t)} \sum_{b=1}^{B} \beta\, e^{-\|\mathbf{x}_u^{(b)}-\mathbf{x}_v^{(b)}\|_2^2}\, \gamma_b(t-t_j),\\
N_{uv}(t) &\sim \text{Hawkes Process}(\lambda_{uv}(t)).
\end{aligned}$$

**The Community Hawkes Independent (CHIP) model**

$$\begin{aligned}
c_u &\sim \mathrm{Categorical}(\pi_1, \ldots, \pi_k), \qquad \forall u \in \mathcal{V}\\
\lambda_{uv}(t) &= \phi_{c_u,c_v} + \sum_{j:t_j\in\mathcal{H}_{v,u}(t)} \alpha_{c_u,c_v}\ \exp\{-(t-t_j)/\beta_{c_u,c_v}\},\\
N_{uv}(t) &\sim \text{Hawkes Process}(\lambda_{uv}(t)).
\end{aligned}$$

**The Hawkes Stochastic Block (Hawkes-SBM) model**

$$\begin{aligned}
c_u &\sim \mathrm{Categorical}(\pi_1, \ldots, \pi_k), \qquad \forall u \in \mathcal{V}\\
\lambda_{k,k'}(t) &= \phi_{k,k'} + \sum_{j:t_j\in\mathcal{H}_{k',k}(t)} \alpha_{k,k'}\ \exp\{-(t-t_j)/\beta_{k,k'}\},\\
N_{k,k'}(t) &\sim \text{Hawkes Process}(\lambda_{k,k'}(t)).
\end{aligned}$$

**The Mutually Exciting Hawkes processes (MHPs) model**

$$\begin{aligned}
\lambda_{uv}(t) &= \phi + \sum_{j:t_j\in\mathcal{H}_{v,u}(t)} \sum_{b=1}^{B} \beta_b\ \gamma_b(t-t_j),\\
N_{uv}(t) &\sim \text{Hawkes Process}(\lambda_{uv}(t)).
\end{aligned}$$

**Poisson process (PPs) model**

$$\begin{aligned}
\lambda_{uv}(t) &= \phi_{uv},\\
N_{uv}(t) &\sim \text{Poisson Process}(\lambda_{uv}(t)).
\end{aligned}$$

## C NUMERICAL SIMULATIONS

In this experiment we use synthetic data to evaluate the performance of the Hawkes-EPM in estimating the kernel parameters. We consider a collection of nodes $|\mathcal{V}| = 100$, and $K = 4$ latent communities. We generated the base rate $\mu_k \sim \text{Uniform}[0, 1]$, and set the kernel parameters $[\alpha_1, \alpha_2, \alpha_3, \alpha_4] = [0.5, 0.88, 1.38, 1.96]$, and $\delta = 0.45$. Via the derived Gibbs sampler, the Hawkes-EPM infers the number of latent communities. As shown in Figure (1), the posterior distributions of the estimated $\{\alpha_k\}$ concentrate toward the true values as the number of observed events is increasing.

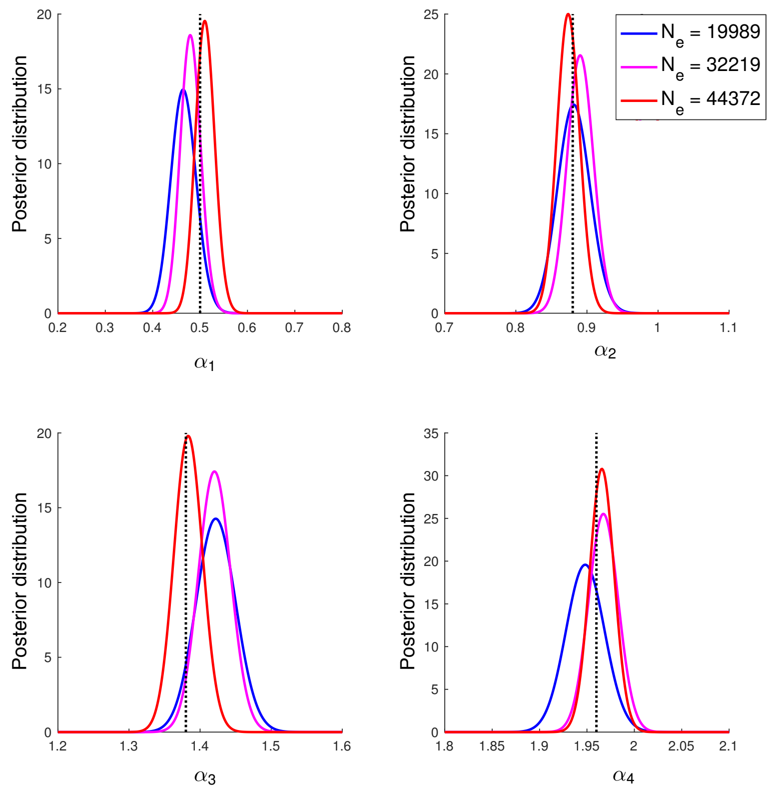


Figure 1: The posterior distribution of the estimated parameters $\{\alpha_k\}$ for the four simulations with the number of events $N_e$. The dashed line indicates the true values of $\{\alpha_k\}$.

# D ADDITIONAL RESULTS

Figures 2 to 4 present the additional plots of the intensities of the interaction events between the nations: Iran (IRN)-USA, Israel (ISR)-Leban (LEB), Israel (ISR)-Palestin(PAL),Iraq (IRQ)-Israel (ISR), Iraq (IRQ)-Kuwait (KUW), Iraq (IRQ)-Saudi Arabi (SAU), USA-Kuwait (KUW), Iraq (IRQ)-Turkey (TUR), United Kingdom (UNK)-Iraq (IRQ).

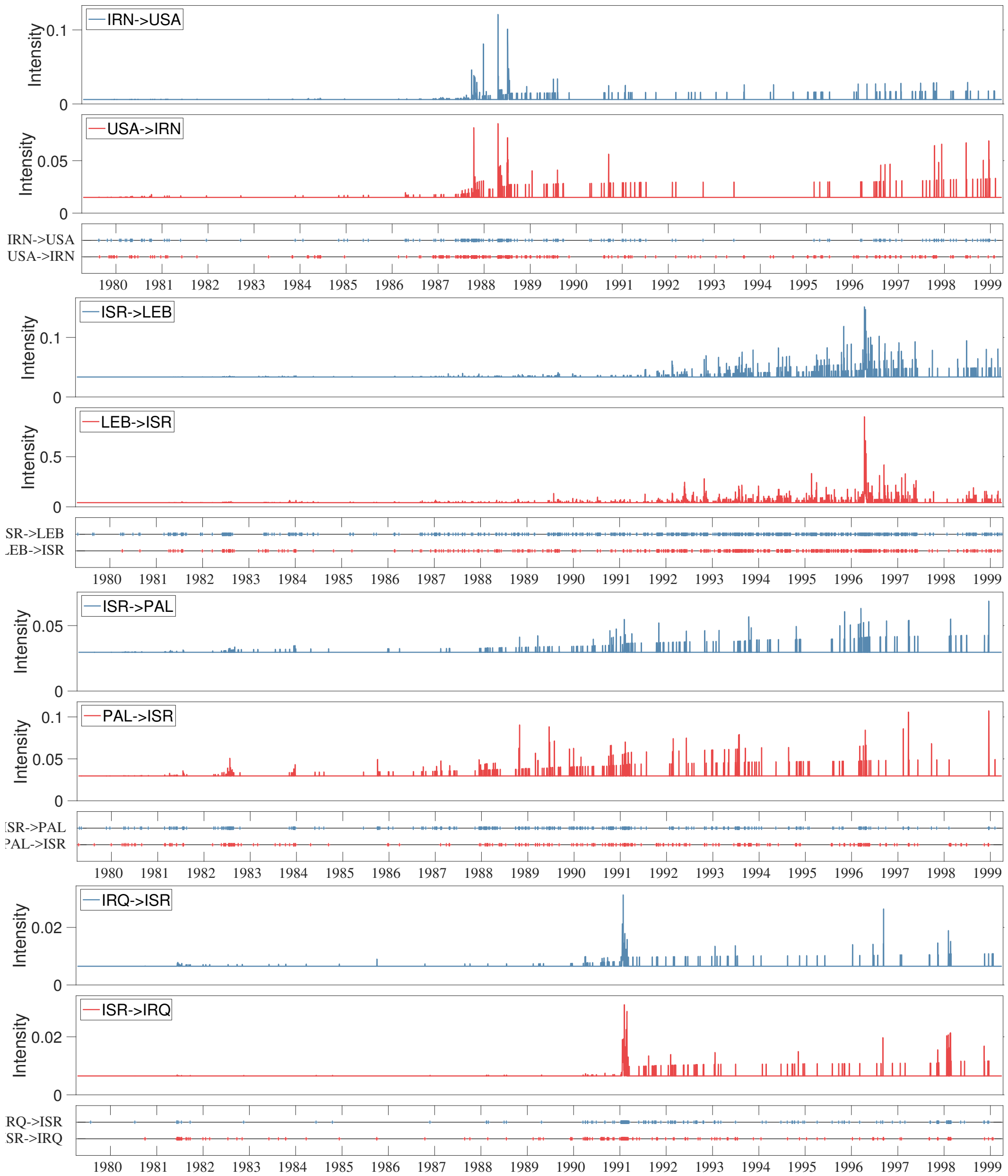


Figure 2: The plots show the intensity of interaction events among nations inferred by the Hawkes-EPM in the Gulf dataset.

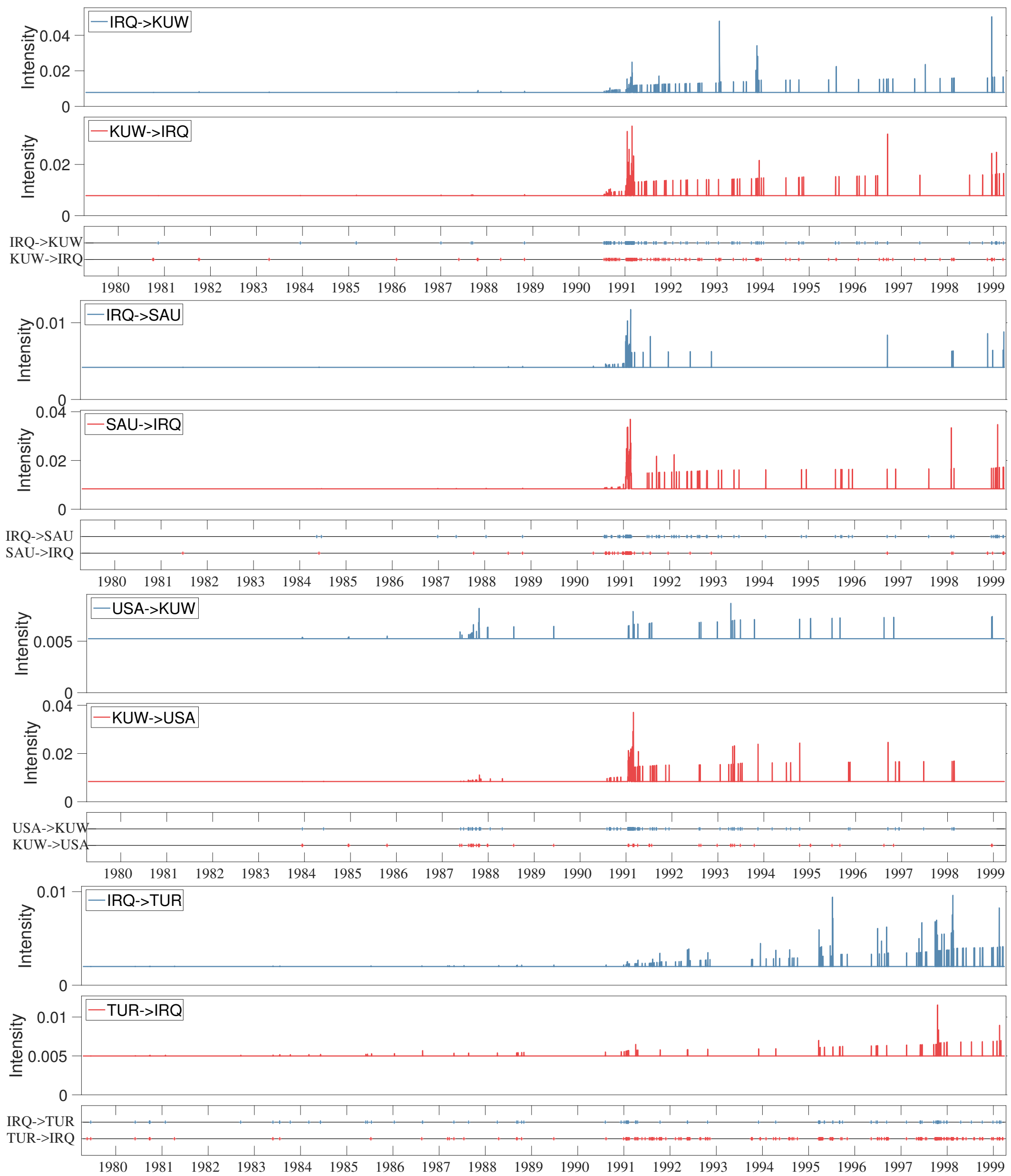


Figure 3: The plots show the intensity of interaction events among nations inferred by the Hawkes-EPM in the Gulf dataset.


Figure 4: The plots show the intensity of interaction events among nations inferred by the Hawkes-EPM in the Gulf dataset.

# References

Polson, N. G. et al. (2013). Bayesian inference for logistic models using Pólya–Gamma latent variables. *Journal of the American Statistical Association*, 108(504):1339–1349.

Yang, J. et al. (2017). Decoupling homophily and reciprocity with latent space network models. In *Proceedings of the Thirty-Third Conference on Uncertainty in Artificial Intelligence*.

Zhou, M. et al. (2012). Lognormal and gamma mixed negative binomial regression. In *Proceedings of the 31st International Conference on Machine Learning*, pages 1343–1350.