Mutually exciting point process graphs for computer network modelling

Francesco Sanna Passino, Nick Heard

Department of Mathematics, Imperial College London

francesco.sanna-passino16@imperial.ac.uk

1. Introduction and motivation

In cyber networks, **relationships between entities**, such as users interacting with computers, or system libraries and the corresponding processes that use them, can provide key insights into **adversary behaviour**. Many cyber attack behaviours create **new links**, initiating previously unobserved relationships between such entities. A **novel** model for **point processes on networks** is proposed to address two fundamental tasks in network security:

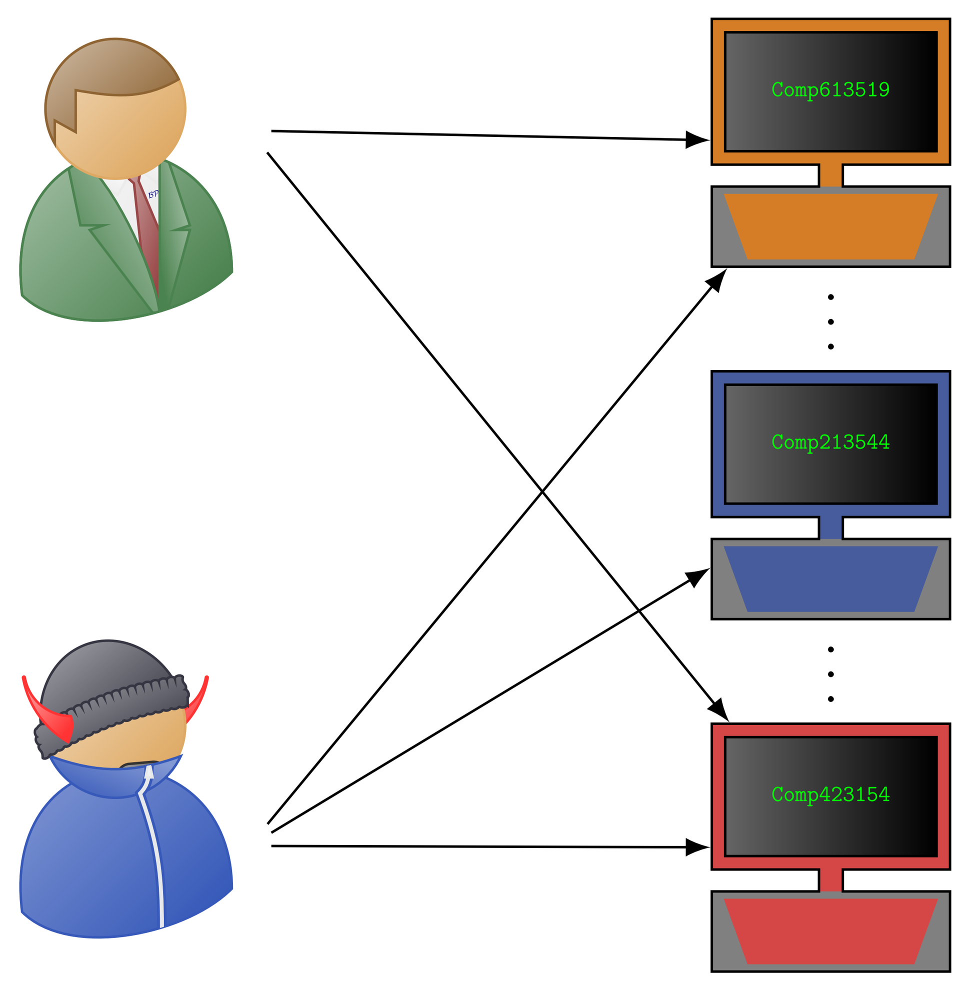
- **network-wide modelling** of event times;
- **anomaly detection** in **new connections**.

2. Computer networks

Computer network data are observed in triplets $(x_1, y_1, t_1), (x_2, y_2, t_2), \dots$, where, for an event (x_i, y_i, t_i):

- x_i and y_i are **marks**, corresponding to the **source** and **destination nodes** from a set of nodes V. For example, x_i could be a user, and y_i an internet server, and the pair (x_i, y_i) forms an **edge**;
- $t_i \in \mathbb{R}_+$ is the **arrival time** of the connection.

The connections on the network can be therefore interpreted as a **point process with dyadic marks**. The main **research objective** is to propose a **network-wide model** for such data.



Acknowledgements

This work is funded by the **Microsoft Security AI** research grant *"Understanding the enterprise: Host-based event prediction for automatic defence in cyber-security"*.

3. Proposed methodology: Mutually Exciting Graphs (MEG)

The **Mutually Exciting Graph (MEG)** uses ideas from mutually exciting processes and latent feature models, combining them into a network-wide point process model framework. A MEG consists of a collection of edge intensity functions $\boldsymbol\lambda(t) = \{\lambda_{ij}(t)\},\ i,j \in V$, of the form:

$$\lambda_{ij}(t) = A_{ij}[\alpha_i(t) + \beta_j(t) + \gamma_{ij}(t)]. \quad (1)$$

- $A_{ij} \in \{0,1\}$ is a **binary constant**, which is 0 if the two nodes i and j are **not** expected to connect, 1 otherwise;
- $\alpha_i(t)$ and $\beta_j(t)$ are the intensity functions corresponding to the **main effects** of the source i and destination j;
- $\gamma_{ij}(t)$ is an **interaction term** between the nodes i and j, parametrised **only** by **node-specific parameters**.

Let $N_{ij}(t)$ be the number of connection events between the nodes i and j before time t, and $N_{i\bullet}(t) = \sum_{j\in V} N_{ij}(t)$, $N_{\bullet j}(t) = \sum_{i\in V} N_{ij}(t)$. Furthermore, denote with $\ell_{i1}, \ell_{i2}, \dots$ the indices $\{k : x_k = i\}$ of the arrival times such that i appears as source node. Also, let $\ell'_{j1}, \ell'_{j2}, \dots$ be the indices $\{k : y_k = j\}$ corresponding to events where j is the destination node. Similarly, let $\ell_{ij1}, \ell_{ij2}, \dots$ be the indices $\{k : x_k = i, y_k = j\}$ of the corresponding events on the edge (i,j). The three functions $\alpha_i(t)$, $\beta_j(t)$ and $\gamma_{ij}(t)$ in (1) are given the following form:

$$\alpha_i(t) = \alpha_i + \sum_{k=N_{i\bullet}(t)-r+1}^{N_{i\bullet}(t)} \omega_i(t - t_{\ell_{ik}}),$$

$$\beta_j(t) = \beta_j + \sum_{k=N_{\bullet j}(t)-r+1}^{N_{\bullet j}(t)} \omega'_j(t - t_{\ell'_{jk}}),$$

$$\gamma_{ij}(t) = \sum_{\ell=1}^{d} \gamma_{i\ell}\gamma'_{j\ell} + \sum_{k=N_{ij}(t)-r+1}^{N_{ij}(t)} \omega_{ij}(t - t_{\ell_{ijk}}),$$

In the above equations:

- $\alpha_i, \beta_j, \gamma_{i\ell}, \gamma'_{j\ell} \in \mathbb{R}_+$ are **baseline intensities**;
- $\omega_i(\cdot), \omega'_j(\cdot)$ and $\omega_{ij}(\cdot) : \mathbb{R}_+ \to \mathbb{R}_+$ are **excitation functions**;
- $r \in \mathbb{N}$ is the **number of past events** that contribute to the intensity. Common choices are $r = 0$ (Poisson process), $r = 1$ (Markov process) and $r \to \infty$ (Hawkes process).

Importantly, $\omega_{ij}(\cdot)$ is parametrised **only** by **node-specific parameters**. The functions $\omega_i(\cdot)$, $\omega'_j(\cdot)$ and $\omega_{ij}(\cdot)$ could be given a **scaled exponential form**, popular for Hawkes processes:

$$\omega_i(t) = \mu_i \exp(-\phi_i t), \qquad \omega'_j(t) = \mu'_j \exp(-\phi'_j t),$$

$$\omega_{ij}(t) = \sum_{\ell=1}^{d} \nu_{i\ell}\nu'_{j\ell} \exp(-\theta_{i\ell}\theta'_{j\ell} t).$$

In $\omega_i(t)$, μ_i could be interpreted as the **jump** in the intensity generated by an observation involving i as source node, whereas ϕ_i expresses **how quickly** the intensity **decays** to the baseline after such an event is observed.

The model **parameters** can be efficiently **learned** using modern **gradient descent algorithms** on the **negative log-likelihood**, for example ***Adam***.

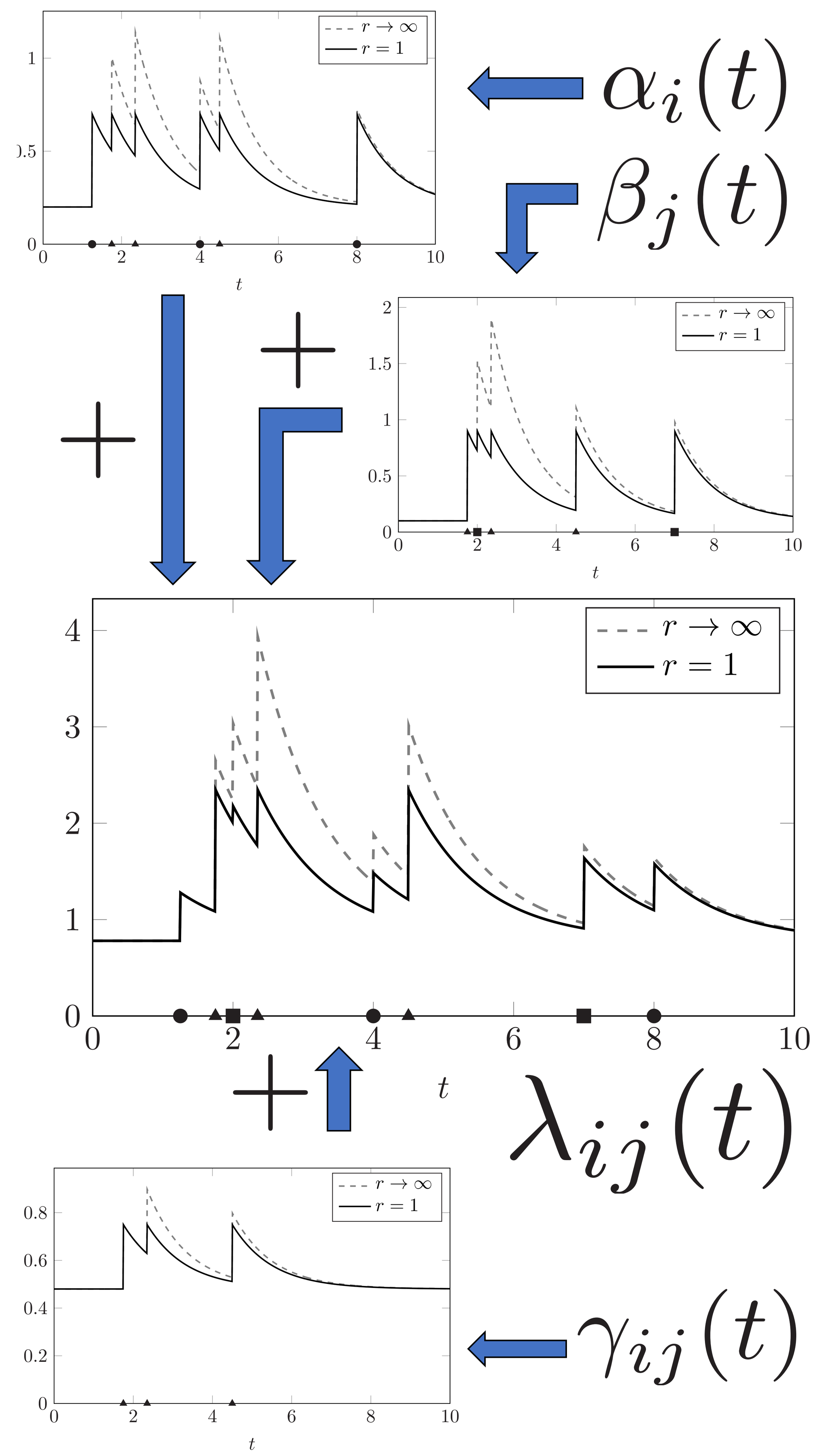


Figure 1: *MEG with scaled exponential excitation.*

4. Results on ICL NetFlow data

NetFlow data are summaries of connections between IP addresses, routinely collected at Imperial College. The MEG model has been fitted on a subset of such data, restricted to **173 clients** hosted within the Department of Mathematics, connecting to **6,083 internet servers**.

	Training set	Test set
Collection period	Jan 20 – Feb 2, 2020	Feb 2 – Feb 9, 2020
Number of arrival times	1,299,372	651,695
Number of edges	115,600	70,408 (40,586 new)

Table 1: *Summary of the subset of ICL NetFlow data used.*

The performance of the MEG models is evaluated using **Kolmogorov-Smirnov scores** on the **test set p-values**. A good value of the score should be **close to 0**, since the p-values should be **uniformly distributed**. The best performance is obtained by a **MEG** with $r = 1$ for main effects and interactions, and $d = 5$, with KS score 0.0738.

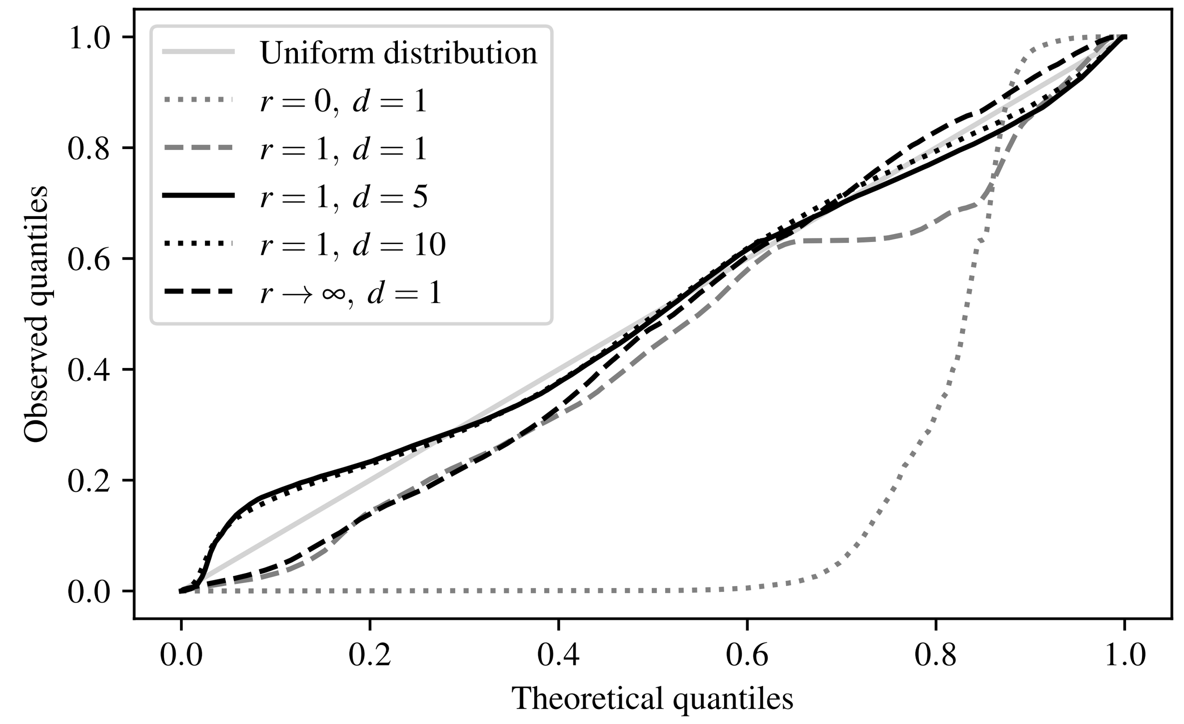


Figure 2: *Q-Q plots for the test p-values obtained from different scaled exponential MEG models, with main effects $\alpha_i(t)$ and $\beta_j(t)$ with $r = 1$, and different parameters for the interaction term $\gamma_{ij}(t)$.*

5. Outcomes and discussion

The **MEG model**, a **network-wide self-exciting model for point processes on graphs**, has been proposed.

- **Scalable**: only node-specific parameters are used;
- **New edge prediction**: MEG provides a **statistically principled** way to score arrival times on **new edges**.

Results on real world computer network data show that:

- **Mutually exciting models** ($r = 1$ and $r \to \infty$) significantly outperform Poisson processes ($r = 0$);
- **Interaction terms** are **essential** to obtain a good predictive performance;
- MEG significantly outperforms **state-of-the-art** methods for point processes on graphs.