# Monitoring a device in a communication network

#### Nick Heard Melissa Turcotte

#### Department of Mathematics, Imperial College London

#### 19 March 2013

Imperial College London

# Monitoring a device

Monitoring a large interaction network over time for anomalous behaviour is computationally very challenging. The task is to build models of normal behaviour in the network, and

then detect anomalous departures from this normal behaviour.

Focus of this work is on monitoring nodes and edges with simple independence assumptions.

If the nodes are computing devices, such as PCs or mobile phones, such models lend themselves to analytics where the analysis can potentially be run on the device itself. For each device, there is an associated multivariate stream of data with diverse characteristics and of differing dimension for different devices.

The aim is to process this multivariate stream into a single time-varying score of surprise.

In both the continuous and discrete time settings, a model is constructed for the full multivariate data stream using conditional independence assumptions.

Surprising recent behaviours against this learnt model of normality are sought.

- 1 VAST 2008 Challenge Data
- 2 Continuous Time Behavioural Modelling
- 3 Discrete Time Behavioural Modelling
- 4 Continuous Time Behavioural Monitoring
- 5 Discrete Time Behavioural Monitoring
- 6 Results for the VAST Data



A publicly available, but synthetic, data set was provided by the VAST 2008 Challenge (http://www.cs.umd.edu/hcil/VASTchallenge08). The data are the mobile phone calls made within a small population of 400 nodes over a ten day period.

Each call event is logged with full details of the meta-data of that call:

- source node
- destination node
- time call started
- time call ended
- cell tower of source caller (geolocation)

 $\rightarrow$  WHO, WHEN & WHERE.

The records should provide critical information about an important social network structure. The aim of the challenge was to detect some anomalous activity from a small subset of the individuals sometime within the ten day period.

From the results of award winning published work on this challenge by Ye et al. (2008), which used a combination of the *PageRank* algorithm (Brin and Page, 1998) and visual analytic methods, there is good reason to suspect that:

- the major anomalous activity occurs on the eighth day
- involves a list of at least eleven individuals





Figure : VAST data by event time and source node

N A Heard & M J M Turcotte (Imperial)

Monitoring a communicating device

19 March 2013 7 / 61



#### Figure : Subgraph of VAST data from malicious actors' contacts

A continuous time view

There are many different ways to represent the status of node *i* using Markov jump processes.

$$i \text{ idle} \xrightarrow{\mu_i} i \text{ connected}$$

 $\mu_i(t)$  is the rate at which node *i* makes connections at time *t*.

There are many different ways to represent the status of node *i* using Markov jump processes.



 $\mu_i(t)$  is the rate at which node *i* makes connections at time *t*.  $\lambda_i(t)$  is the rate at which node *i* terminates connections at time *t*.



 $i \rightsquigarrow$  denotes *i* initiating a connection.  $\rightsquigarrow$  *i* denotes *i* receiving a connection.



Let  $Y_i(t) \in \{0, 1\}$  be an indicator function such that

 $Y_i(t) = 1 \iff i \text{ idle at time } t.$ 

Taking the last model as an example, define  $N_{ij}(t)$  to be the counting process of events  $i \rightsquigarrow j$ .  $N_{ij}(t)$  has intensity function

 $Y_i(t)Y_j(t)\mu_{ij}(t)$ 

Note that the durations of the connections of node *i* (to any other node) and node *j* (to any other node) both act as a censoring mechanism for the process  $N_{ij}(t)$ .

If connection durations are also of inferential interest, let  $\bar{N}_{ij}(t)$  be the counting process of  $i \rightsquigarrow j$  connection terminations. Also let  $Y_{ij}(t) \in \{0, 1\}$  be an indicator function such that

$$Y_{ij}(t) = 1 \iff i \text{ connected to } j \text{ at time } t.$$

 $\bar{N}_{ij}(t)$  has intensity function

 $Y_{ij}(t)\lambda_{ij}(t)$ 

Note that the pair of processes  $N_{ij}(t)$ ,  $\bar{N}_{ij}(t)$  are heavily dependent upon one another; one always has zero intensity. However, their p-values (considered later) are asymptotically independent of one another.

# Seasonality

Fast, tractable conjugate Bayesian inference is available with gamma priors if the state transition intensities  $\lambda_{ij}(t)$  and  $\mu_{ij}(t)$  are assumed constant.

 $\rightarrow$  the inherent seasonality in the processes does not admit constant intensities for representing normal behaviour.

For example, looking at a daily level, it is likely there will be variability in connectivity between the night time, the day time, evening, and so on.

Let *S* be a seasonal period over which the processes are expected to show repetitive intensity patterns; for example, *S* might be the length of one day.

The process  $N_{ij}(t)$  will be assumed to be the censored counting process (implied by the state space diagrams) of events arising from an inhomogeneous Poisson process with intensity function

 $\mu_{ij}(t) = \mu^{ij} m_{ij}(t \bmod S)$ 

where  $m_{ij} : [0, S] \to \mathbb{R}^+$  is a probability density function for the time within the season at which a **single** connection would be made. So the intensity function of the process  $N_{ij}(t)$  for normal behaviour is

$$\mu^{ij} Y_i(t) Y_j(t) m_{ij}(t \bmod S)$$

$$(2.1)$$

Similarly,  $\bar{N}_{ij}(t)$  will have intensity

$$\lambda^{ij} Y_{ij}(t) l_{ij}(t \bmod S)$$
(2.2)

where  $l_{ij} : [0, S] \to \mathbb{R}^+$  is the probability density function of a single connection ending time for the edge.

For (approximately) known density functions  $l_{ij}$ ,  $m_{ij}$ , this construction then admits conjugate Bayesian inference when

 $\lambda^{ij}, \mu^{ij} \sim \Gamma(a, b).$ 

The density functions  $l_{ij}$ ,  $m_{ij}$  for capturing seasonality must be learnt from training data, and then periodically updated.

If some edges have sparse data, then it can be reasonable to assume that for  $j \neq k$ ,

$$l_{ij} = l_{ik} = l_i,$$
$$m_{ii} = m_{ik} = m_i$$

and possibly even  $l_i = m_i$ . This also reduces the computational resource required.

### A changepoint model for $l_{ij}$ , $m_{ij}$

Assume a piecewise constant density on [0, S), with  $k_S$  changepoints  $\sigma_{1:k_S}$  ordered such that  $0 = \sigma_0 < \sigma_1 < \ldots < \sigma_{k_S} < \sigma_{k_S+1} = S$ . Changepoints are assumed to follow a Poisson process with rate  $\nu_S$ ,

$$p(k_S,\sigma_{1:k_S})=\nu_S^{k_S}e^{-\nu_S S}$$

Conditional on  $(k_S, \sigma_{1:k_S})$ , let  $\theta_j$  be the probability of an observation falling in the *j*<sup>th</sup> segment,  $j = 1, 2, ..., k_S + 1$ . Together the changepoints and these probabilities specify a piecewise constant density, say

$$m_{ij}(s) = \sum_{j=1}^{k+1} \mathbb{I}_{[\sigma_{j-1},\sigma_j)}(s) \frac{\theta_j}{\sigma_j - \sigma_{j-1}}, \ s \in [0,S).$$

Straightforward conjugate inference can be completed by specifying

$$[\theta_{1:k_S+1}|(k_S,\sigma_{1:k_S})] \sim \text{Dirichlet}(\alpha_1,\alpha_2,\ldots,\alpha_{k_S+1}),$$

where  $\alpha_j = \alpha\{[\sigma_{j-1}, \sigma_j)\}$  and  $\alpha(\cdot)$  is a base measure on [0, S); the default choice being Lebesgue measure,

$$\alpha\{[\sigma_{j-1},\sigma_j)\} = \alpha \frac{(\sigma_j - \sigma_{j-1})}{S}.$$

- Wrap each observation into the season [0, S):  $\rightarrow$  replace *t* with *t* mod *S*.
- Let  $n_j$  be the number of wrapped observations falling in the  $j^{\text{th}}$  segment defined by the changepoints, and  $n = \sum_{j=1}^{k_S+1} n_j$ .

Care has to be taken not to be double counting the censorship induced by the rival processes. Equations (2.1) and (2.2) already account for the fact that node i is not available to make a new connection whilst it is still engaged in an existing connection. So the density estimates should not reflect this censoring.

Strictly, this would require treating each seasonal period of data as a sample from a truncated version of  $m_{ij}(s)$ , but this breaks the conjugacy of the Dirichlet model. So instead an approximate solution is sought.

Focusing on the connections made by node *i*, for  $s \in [0, S]$ , after observing  $n_S$  seasons let

$$\tilde{Y}_i(s) = \sum_{j=0}^{n_S} Y_i(s+jS)$$

be the number of seasons in which the node was being observed and available to make a connection at time *s*.



Figure :  $\tilde{Y}_1(s)$  for the VAST data

#### Define

$$a_j = \int_{s=\sigma_{j-1}}^{\sigma_j} \tilde{Y}_i(s) lpha(ds),$$
  
 $s_j = \int_{s=\sigma_{j-1}}^{\sigma_j} \tilde{Y}_i(s) ds$ 

to be, respectively, the base measure and Lebesgue measure of the total observation time of node i in the j<sup>th</sup> segment, and

$$a = \sum_{j=1}^{k_S+1} a_j$$

to be the base measure of the overall total observation time for node *i*.

Then approximate inference can proceed by defining

$$q_j = \frac{s_j}{n_S(\sigma_j - \sigma_{j-1})},$$
$$\theta'_j = \frac{q_j \theta_j}{\sum_{i=1}^{k+1} q_i \theta_i}$$

as the actual probabilities of observing the *j*<sup>th</sup> segment, and then performing conjugate inference assuming

$$[\theta'_{1:k_S+1}|(k_S,\sigma_{1:k_S})] \sim \text{Dirichlet}(a_1,a_2,\ldots,a_{k_S+1}).$$

An estimate of  $\theta'$  is then transformed back into probabilities  $\theta$  which account for censoring via

$$heta_j = rac{ heta_j'/q_j}{\sum_{i=1}^{k+1} heta_i'/q_i}.$$

Reversible jump MCMC sampling changepoints from the posterior distribution is trivial via

$$p(k_S, \sigma_{1:k_S}|\mathcal{D}) \propto \nu_S^{k_S} \prod_{j=1}^{k_S+1} \frac{\Gamma(a_j+n_j)}{\Gamma(a_j)s_j^{n_j}}.$$

The MAP number of changepoints is first obtained,

$$k_S^* = \operatorname*{argmax}_{k_S} p(k_S | \mathcal{D})$$

and then conditional on  $k_S = k_S^*$ , the MAP changepoints are obtained,

$$\sigma_{1:k_S}^* = \operatorname*{argmax}_{\sigma_{1:k_S^*}} p(k_S^*, \sigma_{1:k_S^*} | \mathcal{D}).$$

The transformed posterior mean heights for the piecewise constant probability density function corresponding to these changepoints are given by

$$m_j^* \propto \frac{a_j + n_j}{(a+n)s_j}, \quad j = 1, 2..., k_S + 1.$$

# Fitted density $m_3(s)$ for node 3 from the VAST data



Figure : Estimated  $m_3(s)$  for the VAST data

A discrete time view

### Node activity status

Connection event data can also be modelled in discrete time, with connection counts aggregated into, say, one minute intervals.

Let  $Z_i(t)$  be the number of connections for node *i* at the *t*<sup>th</sup> time point. For a binary perspective, again let  $Y_i(t) \in \{0, 1\}$  be the indicator variable for whether node *i* is idle at time *t*; so in discrete time,

$$Y_i(t) = 0 \iff Z_i(t) > 0.$$



### Node activity status

Connection event data can also be modelled in discrete time, with connection counts aggregated into, say, one minute intervals.

Let  $Z_i(t)$  be the number of connections for node *i* at the *t*<sup>th</sup> time point. For a binary perspective, again let  $Y_i(t) \in \{0, 1\}$  be the indicator variable for whether node *i* is idle at time *t*; so in discrete time,



A Markov chain for connectivity status has transition probability matrix

$$P = \left( egin{array}{cc} \phi^i & 1 - \phi^i \ \psi^i & 1 - \psi^i \end{array} 
ight)$$

where

$$\phi^{i} = \mathbb{P}(Y_{i}(t) = 0 | Y_{i}(t-1) = 0)$$
  
$$\psi^{i} = \mathbb{P}(Y_{i}(t) = 0 | Y_{i}(t-1) = 1)$$

This model assumes connection durations to follow a Geometric( $\phi^i$ ) distribution rather than the exponential distribution from continuous time, although here no event distinction is made if a connection is terminated and a new one begins within the same discrete time period.

... An analogous seasonal changepoint model can be constructed.

Continuous time monitoring

Recall equation (2.1) for the intensity of the process  $N_{ij}(t)$  for normal behaviour,

$$\mu^{ij} Y_i(t) Y_j(t) m_{ij}(t \bmod S)$$

The seasonality will be considered stable under the null hypothesis, and local changes in the overall level of connectivity from normal behaviour will act as changepoints in  $\mu^{ij}$ .

For a sequence of monitoring times  $t_1 < t_2 < ...$ , inference will be made about the process based on observation over  $[t_0, t_n]$ .

- The changepoints  $\tau_{1:k_n} = (\tau_1, \dots, \tau_{k_n})$  arrive as a homogeneous Poisson process with intensity  $\nu$
- The intensities have conjugate priors,

$$\mu_{0:k_n}^{ij} = (\mu_0^{ij}, \ldots, \mu_{k_n}^{ij}) \stackrel{\text{iid}}{\sim} \Gamma(\alpha, \beta),$$

and so these can be integrated without any estimation.

So the changepoint posterior distribution is known up to proportionality,

$$\pi_{[t_0,t_n]}(\tau_{1:k_n},k_n) \propto \gamma_{[t_0,t_n]}(\tau_{1:k_n},k_n) = \nu \mathrm{e}^{-\nu t} \prod_{k=0}^{k_n} \frac{\beta^{\alpha}}{\Gamma(\alpha)} \frac{\Gamma(\alpha+r_k)}{(\beta+y_k)^{r_k+\alpha}}$$

where  $\tau_0 = t_0$ ,  $\tau_{k+1} = t_n$  and  $r_k$  is the number of observed connections in the  $k^{\text{th}}$  segment and

$$y_k = \int_{t=\tau_k}^{\tau_{k+1}} Y_i(t) Y_j(t) m_{ij}(t \bmod S) dt$$

is the seasonally rescaled total observation time in the  $k^{\text{th}}$  segment.

- $(\tau_{1:k_1}, k_1)^{(i)}$  are easily sampled from  $\pi_{[t_0, t_1]}(\tau_{1:k_1}, k_1)$  by RJMCMC.
- Subsequently, sequential Monte Carlo (SMC) sampling can exploit the similarity of  $\pi_{[t_0,t_{n-1}]}$  and  $\pi_{[t_0,t_n]}$ . Turcotte and Heard (2013).

# A fast SMC algorithm for changepoints

At  $t_{n-1}$ , let  $\{(\tau_{1:k_{n-1}}, k_{n-1})^{(i)}, w_{n-1}^{(i)}\}_{i=1}^N$  be a sample of particles and importance weights for the target  $\pi_{[t_0, t_{n-1}]}$ .

Define  $t_{n-1}^*$  to be the time of the most recent changepoint before  $t_{n-1}$ . The posterior expectation of  $t_{n-1}^*$  wrt  $\pi_{[t_0,t_{n-1}]}$  can be calculated using the particle approximation

$$\mathbb{E}_{\pi_{[t_0,t_{n-1}]}}[t_{n-1}^*] \approx \sum_{i=1}^N \tau_{k_{n-1}}^{(i)} w_{n-1}^{(i)}.$$

Additional changepoints are then proposed for the update interval  $(t_{n-1}, t_n]$ , using RJMCMC for the posterior restricted to the data from  $(t_{n-1}^*, t_n]$ . Note that changepoints are only sampled from  $(t_{n-1}, t_n]$ .

Note that this is equivalent to proposing additional changepoints from the posterior  $\pi_{[t_{n-1},t_n]}$  with a revised prior distribution for the first intensity level of

$$\Gamma(\alpha + d_{n-1}^*, \beta + t_{n-1} - t_{n-1}^*),$$

where  $d_{n-1}^*$  is the number of events observed within  $(t_{n-1}^*, t_{n-1}]$ .

The proposed changepoint samples for the update interval  $(t_{n-1}, t_n]$  are then appended to the existing samples.

- The new sample is permuted to break correlation.
- If an update interval has low probability of containing changepoints, then fewer samples can be drawn for that update.

## Time since last changepoint

As a measure of anomaly at time t, let g(t) be the time that has passed since the last changepoint,

$$g(t) = t - \tau_{i^*},$$
  
$$i^* = \max_i \{\tau_i \le t\}.$$

Small values of g(t) correspond to recent change and therefore anomalous behaviour.

Note that the prior expectation and variance of g(t) are both increasing with t, and therefore it is preferable to work with a standardised alternative

$$h(t) = \frac{g(t) - \mathbb{E}[g(t)]}{\sqrt{\operatorname{Var}[g(t)]}}.$$

For monitoring, there are several possibilities, including:

- The posterior expectation of *h*(*t*) this should take highly negative values for an anomaly.
- The posterior probability ℙ(h(t) < 0) > this should be high for an anomaly.

The latter has the advantage that it is easily calibrated, and anomalies can be flagged whenever

 $\mathbb{P}(h(t) < 0) > \alpha$ 

for some  $\alpha$ , say 0.95.

Discrete time monitoring

Suppose an inhomogeneous Poisson process has recently been observed over  $(t_{n-1}, t_n]$ , producing two pieces of data:

- a total number of events  $N(t_n) N(t_{n-1}) = k$ ;
- event times  $t_{n-1} = t^{(0)} < t^{(1)} < ... < t^{(k)} < t_n$  and inter-arrival times  $x_1 = t^{(1)} t_{n-1}, x_2 = t^{(2)} t^{(1)}, x_k = t^{(k)} t^{(k-1)}$  and perhaps a right-censored arrival time  $x_{k+1} = t_n t^{(k)}$ .

Respectively there are two natural p-values to consider (and corresponding lower tail and two-sided analogues):

2  $s_{x_1}, s_{x_2}, \ldots, s_{x_k}, s^*_{x_{k+1}}$  combined via Fisher's method. Sverdrup's (unpublished) observations on transition intensities imply that the former would be more powerful when the intensity of connections is low, otherwise the latter would have more power.

### Control charts

For a collection of independent p-values  $p^1, \ldots, p^k$ , Fisher's method combines these separate measures of surprise into a single score

$$X^2 = -2\sum_{i=1}^k \log p^i.$$

When the null hypotheses are correct,  $X^2 \sim \chi^2_{2k}$ .

At monitoring time point  $t_n$ , the aim is to combine independent uniform p-values obtained for the most recently observed data from each aspect of the multivariate data stream using Fisher's method, to give a single measure of surprise  $p_n$ .

### Stouffer's Z-score

Combined p-values  $\{p_n\} \stackrel{\text{iid}}{\sim} [0, 1]$  under the null hypothesis of normal behaviour, so if

$$Z_n=\Phi^{-1}(1-p_n),$$

then  $\{Z_n\} \stackrel{\text{iid}}{\sim} N(0,1)$ .

Now outlying (small p-value) behaviour at time  $t_n$  will correspond to a large value of  $Z_n$ .

Finally, to accumulate evidence of anomalous behaviour over time, the proposed method is to run an exponentially weighted moving average (*EWMA*) chart  $\{S_n\}$  on the Z-scores  $\{Z_n\}$ ,

$$S_0 = 0,$$
  
 $S_n = (1 - w)S_{n-1} + wZ_n, \ n \ge 1,$ 

with the dual benefits that more recent values carry highest weight but recent surprise over successive intervals can also be accumulated. The tunable parameter  $w \in [0, 1]$  controls the level of significance placed on the most recent Z-score.

This model has well understood boundaries under the null,

$$L\sqrt{\frac{w}{2-w}[1-(1-w)^{2n}]}.$$

# Adjusting discrete p-values

The discrete time analytics proposed here rely on p-values from either binary or count random variables.

For example, for the discrete time model which takes a binary view of activity status, for computational tractability p-values need to be calculated for each calculated binary observation of activity status and then combined using a method such as Fisher's. The crude discrete p-values can only take two values, and so are far from uniform. Let *X* be a discrete random variable on  $\mathbb{N} = \{0, 1, 2, ...\}$ . Denote the probability mass function (pmf) and survivor function respectively as

$$p_x = \mathbb{P}(X = x)$$
$$s_x = \mathbb{P}(X \ge x) = 1 - \sum_{j=0}^{x-1} p_j$$

• *s<sub>x</sub>* are the upper tail p-values,

$$\mathbb{P}(s_X = s_i) = p_i, \quad i = 0, 1, 2, \dots$$

- Note that  $s_0 \equiv 1$  and  $\lim_{x\to\infty} s_x = 0$ .
- W.l.o.g. can assume  $p_0 > 0$ .
- e.g. for binary variables, could further assume *p*<sub>0</sub> ≥ *p*<sub>1</sub> ≥ .. if that sort of p-value preferred.

A problem for this work is that p-values for discrete random variables are discrete and therefore not U[0, 1] random variables. Instead they are only approximately uniform in the following sense: If

$$X \sim p_x,$$
  
and  $U|X = x \sim U(s_{x+1}, s_x]$ 
$$\implies U \sim U[0, 1]$$

Proof: For  $u \in [0, 1]$ , *U* has density function

$$f(u) = \sum_{x} p_{x} f(u|x) = \sum_{x} (s_{x} - s_{x+1}) \frac{\mathbb{I}_{(s_{x+1}, s_{x}]}(u)}{s_{x} - s_{x+1}} = \sum_{x} \mathbb{I}_{(s_{x+1}, s_{x}]}(u) = \mathbb{I}_{[0,1]}(u)$$

So in fact, discrete p-values are stochastically larger than U[0,1] random variables. If  $U \sim U[0,1]$ ,

$$\mathbb{P}(U < s_X) = \sum_x p_x \mathbb{P}(U < s_x) = \sum_x p_x s_x = \sum_x p_x \sum_{j \ge x} p_j = \frac{1 + \sum_x p_x^2}{2} > \frac{1}{2}.$$

- Unadjusted discrete p-values are too big!
- In particular,  $\mathbb{E}(s_X) > \frac{1}{2}$ .
- The same results hold for lower tail p-values.

0

Really, discrete p-values should be regarded as an interval censored observation of a truly uniform p-value.

- cf. the p-value of an interval censored observation of a continuous random variable.
- A discrete observation of X = x corresponds to the censored observation of a p-value

 $s \sim \mathrm{U}(s_{x+1},s_x].$ 

# A Monte Carlo adjustment

- A deterministic correction can be made, e.g.  $s = \frac{s_{\chi+1} + s_{\chi}}{2}$ .
- Alternatively, each observed p-value can be preserved as a  $U(s_{x+1}, s_x]$  random variable.

Suppose several independently observed p-values  $s^{(i)}$  are going to be combined, typically arising from different distributions. If

 $s^{(i)} \sim \mathrm{U}(a_i, b_i]$ 

then the expected value of Fisher's score can be taken with respect to their joint density

$$\prod_i \frac{\mathbb{I}_{(a_i,b_i]}(s^{(i)})}{b_i-a_i}.$$

Monte Carlo estimation of expectations arising from this joint distribution are trivial to perform but carry an increased computational cost.

#### Some results for the VAST data

The following graphs show control charts or changepoint curves for the different model and analytic combinations:

- Continuous time model (Markov jump process), discrete time analytic S<sub>t</sub>.
- Continuous time model(Markov jump process), continuous time analytic 𝔼[*h*(*t*)] or 𝒫(*h*(*t*) < 0).</p>
- **3** Discrete time model (Markov chain), discrete time analytic  $S_t$ .

Each line represents one node from the network.

- If present, a green line shows a threshold for decision making.
- Curves of malicious actors in the data set are coloured
  - blue if no threshold has been used or if they were not detected against the threshold
  - otherwise they are coloured red to indicate that they were detected.
- False detections are coloured purple.



Figure : VAST data control charts: continuous time model for undirected node event times



Figure : VAST data control charts: continuous time model for outgoing event times for each node



Figure : VAST data control charts: continuous time model for incoming event times for each node



Figure : VAST data control charts: continuous time model for incoming event times and Markov model for corresponding incoming caller for each node



Figure : VAST data control charts: continuous time model for incoming event times and Markov model for corresponding call towers for each node



Figure : VAST data control charts: continuous time models for incoming and outgoing event times for each node



Figure : VAST data  $\mathbb{E}[h(t)]$ : continuous time model for incoming event times for each node



Figure : VAST data  $\mathbb{P}(h(t) < 0)$ : continuous time model for incoming event times for each node

Continuous time modelling gives more accurate inference and is computationally faster than discrete time modelling, which also requires an arbitrary discretisation of the time domain.

In contrast, continuous time monitoring and discrete time monitoring give very similar inference, and it is discrete time monitoring that has the greater computational simplicity (with no requirement for costly MCMC or SMC simulation).

# Summary of VAST data results

The majority of the signal for detecting the malicious actors was in the change in pattern of their incoming calls. The outgoing calls held almost no information. This finding stood up for all of the different analytics.

Beyond this, there was no further information to be extracted from the identity of the incoming callers, or the cell towers used by those callers.

Splitting the data stream into separate process for each edge led to more noise rather than more signal, the information here was in the aggregated call pattern.

This is only synthetic data, although it is interesting to be able to gain insight into how the data were generated.

- Brin, S. and Page, L. (1998) The anatomy of a large-scale hypertextual Web search engine. *Computer Networks and ISDN Systems*, **30**, 107–117.
- Ye, Q., Zhu, T., Hu, D., Wu, B., Du, N. and Wang, B. (2008) Cell Phone Mini Challenge Award: Social Network Accuracy - Exploring temporal communication in mobile call graphs. In *Proceedings of IEEE VAST Symposium*, 207–208. Piscataway, NJ: IEEE.