Network anomaly detection: the state of the art

Patrick Rubin-Delanchy
patrick.rubin-delanchy@bristol.ac.uk

School of Mathematics, University of Bristol

December 17, 2018

Network modelling

The methods you have learnt to model dynamic networks may seem unsatisfactory. What about:

1. changepoints
2. community structure
3. information flow
4. exploiting the typically rich additional data available which could include covariates at event (e.g. port number, or text), edge, node (e.g. geography) levels.

Network modelling

The methods you have learnt to model dynamic networks may seem unsatisfactory. What about:

1. changepoints
2. community structure
3. information flow
4. exploiting the typically rich additional data available which could include covariates at event (e.g. port number, or text), edge, node (e.g. geography) levels.

Network modelling

The methods you have learnt to model dynamic networks may seem unsatisfactory. What about:

1. changepoints
2. community structure
3. information flow
4. exploiting the typically rich additional data available which could include covariates at event (e.g. port number, or text), edge, node (e.g. geography) levels.

Network modelling

The methods you have learnt to model dynamic networks may seem unsatisfactory. What about:

1. changepoints
2. community structure
3. information flow
4. exploiting the typically rich additional data available which could include covariates at event (e.g. port number, or text), edge, node (e.g. geography) levels.

Los Alamos National Laboratory data — graph

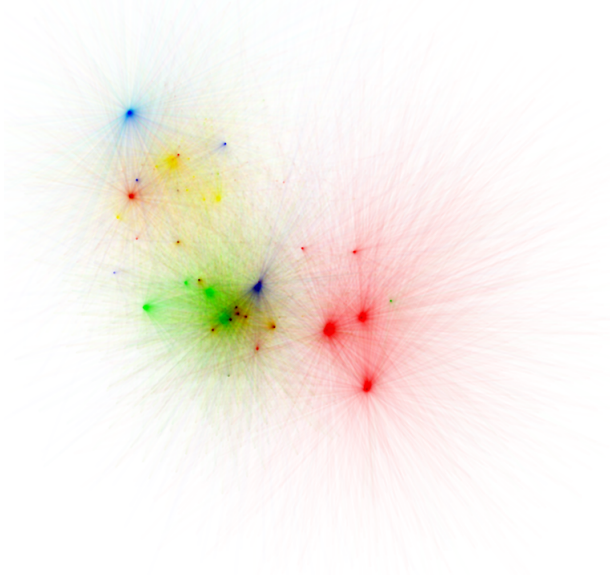


Figure: NetFlow (full graph), Los Alamos National Laboratory (LANL) network (Kent, 2016)

Los Alamos National Laboratory data — time series

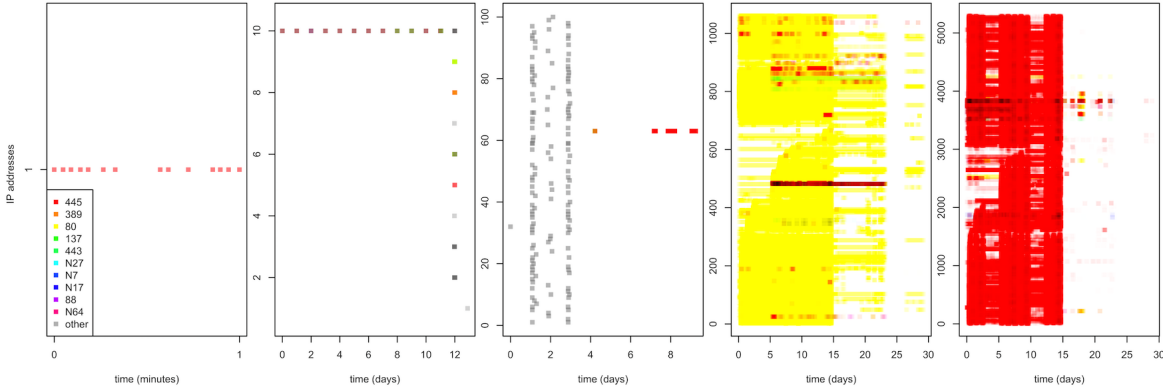


Figure: NetFlow for 5 IP addresses (full collection), LANL network

On the network front, there are the beginnings of a unified theory on statistical modelling of graphs, including:

- The graphon model
- (Generalised) random dot product graphs and their connection to spectral embedding
- The stochastic block model

Aldous-Hoover theorem

Let G be an infinite undirected node exchangeable graph.

Then each edge $i \leftrightarrow j$ is a conditionally independent Bernoulli variable with success probability $f(U_i, U_j)$ where

- $U_1, U_2, \ldots$ are independent uniform random variables on $[0, 1]$.
- f is a symmetric function from $[0, 1]^2 \to [0, 1]$ (called a graphon).

Generalised random dot product graph model

Let $\mathbf{I}_{p,q} = \text{diag}(1, \ldots, 1, -1, \ldots, -1)$, with p ones followed by q minus ones on its diagonal.

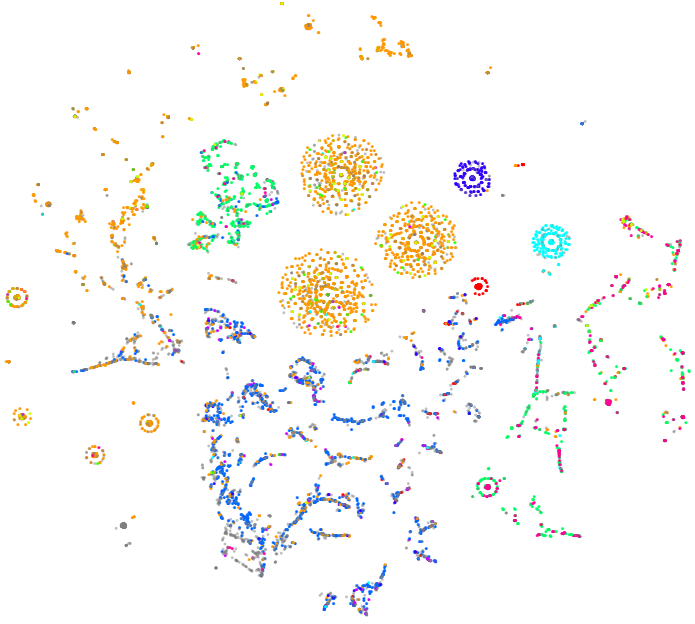
Let $X_1, \ldots, X_n \overset{i.i.d.}{\sim} F$, for some distribution F supported on a valid set $\chi \subset \mathbb{R}^d$.

Then let each edge $i \leftrightarrow j$ be a conditionally independent Bernoulli variable with success probability $X_i^T \mathbf{I}_{p,q} X_j$.

Estimate via spectral embedding:

1. adjacency matrix $\mathbf{A}$
2. eigendecomposition $\mathbf{A} = \mathbf{U}\mathbf{S}\mathbf{U}^\top$, eigenvalues ordered by magnitude
3. $\mathbf{X} = \left[\mathbf{U}|\mathbf{S}|^{1/2}\right]_{1:n,1:d}$
4. $\hat{X}_i = i$th row of $\mathbf{X}$

Spectral embedding ($d = 10$) & t-SNE



The stochastic block model

Let $\mathbf{B} \in [0,1]^{K \times K}$ be a symmetric matrix and $\pi_1, \ldots, \pi_K$ satisfying $\sum_{k=1}^{K} \pi_k = 1$, $\pi_k \geq 0$.

Assign each node to a cluster, $Z_i \overset{i.i.d.}{\sim} \text{Categorical}(\pi_1, \ldots, \pi_K)$.

Then let each edge $i \leftrightarrow j$ be a conditionally independent Bernoulli variable with success probability $\mathbf{B}_{Z_i Z_j}$.

The stochastic block model

Let $\mathbf{B} \in [0,1]^{K \times K}$ be a symmetric matrix and $\pi_1, \ldots, \pi_K$ satisfying $\sum_{k=1}^{K} \pi_k = 1$, $\pi_k \geq 0$.

Assign each node to a cluster, $Z_i \overset{i.i.d.}{\sim} \text{Categorical}(\pi_1, \ldots, \pi_K)$.

Then let each edge $i \leftrightarrow j$ be a conditionally independent Bernoulli variable with success probability $\mathbf{B}_{Z_i Z_j}$.

Theorem

A stochastic block model is a special case of the generalised random dot product graph model, which is itself a special case of the graphon model.

While there has been important recent progress on graph modelling, dynamic networks such as those observed in cyber-security applications are much more complex.

Of course, alongside this more theoretical research there have been momentous successes in machine-learning with complex data, particularly with images and text.

However, an important current limitation of these methods is providing probabilistic predictions and measures of uncertainty, so that the problem anomaly detection (which requires both) is for most modern applications effectively unsolved.