New edge prediction and anomaly-detection in dynamic networks

Silvia Metelli

10 April, 2019

Centre de recherche epidémiologie et statistique Sorbonne Paris Cité

The Alan Turing Institute

Imperial College London

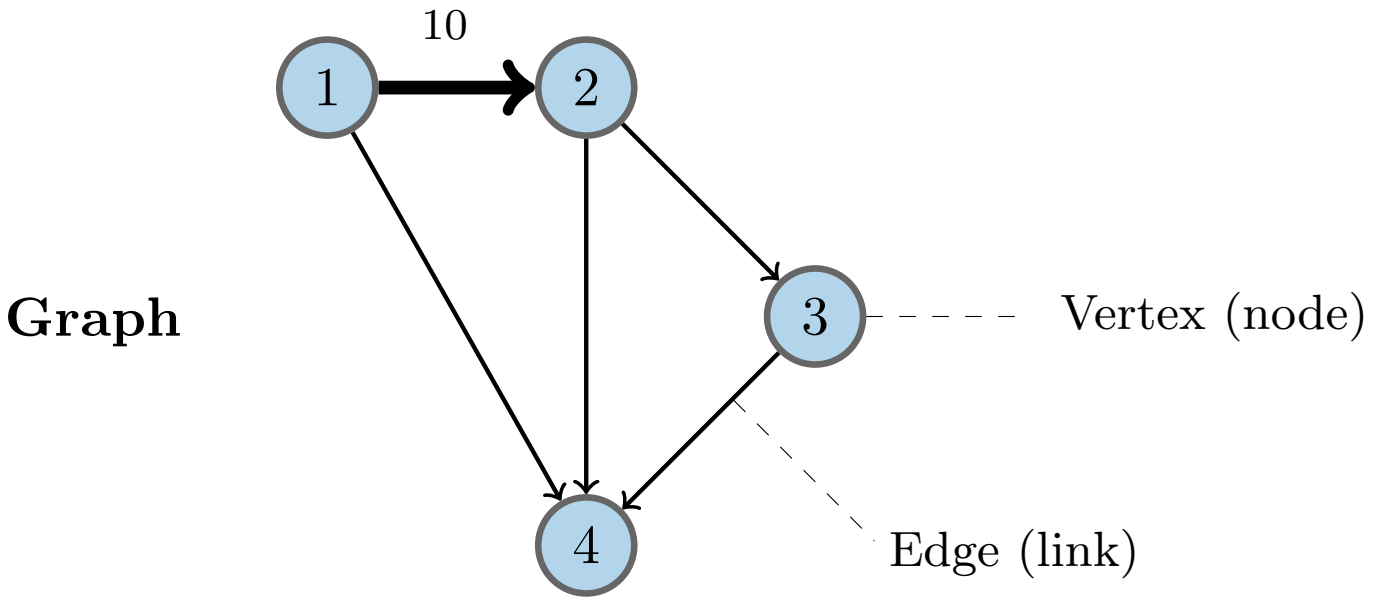
Collaborators

- Nick Heard (Imperial College London & Heilbronn Institute for Mathematical Research)
- Niall Adams (Imperial College London & Heilbronn Institute for Mathematical Research)
- Kaushik Jana (The Alan Turing Institute)
- Simone Claudiani (Imperial College London, Department of Haematology NHS Trust)

Overview of this talk

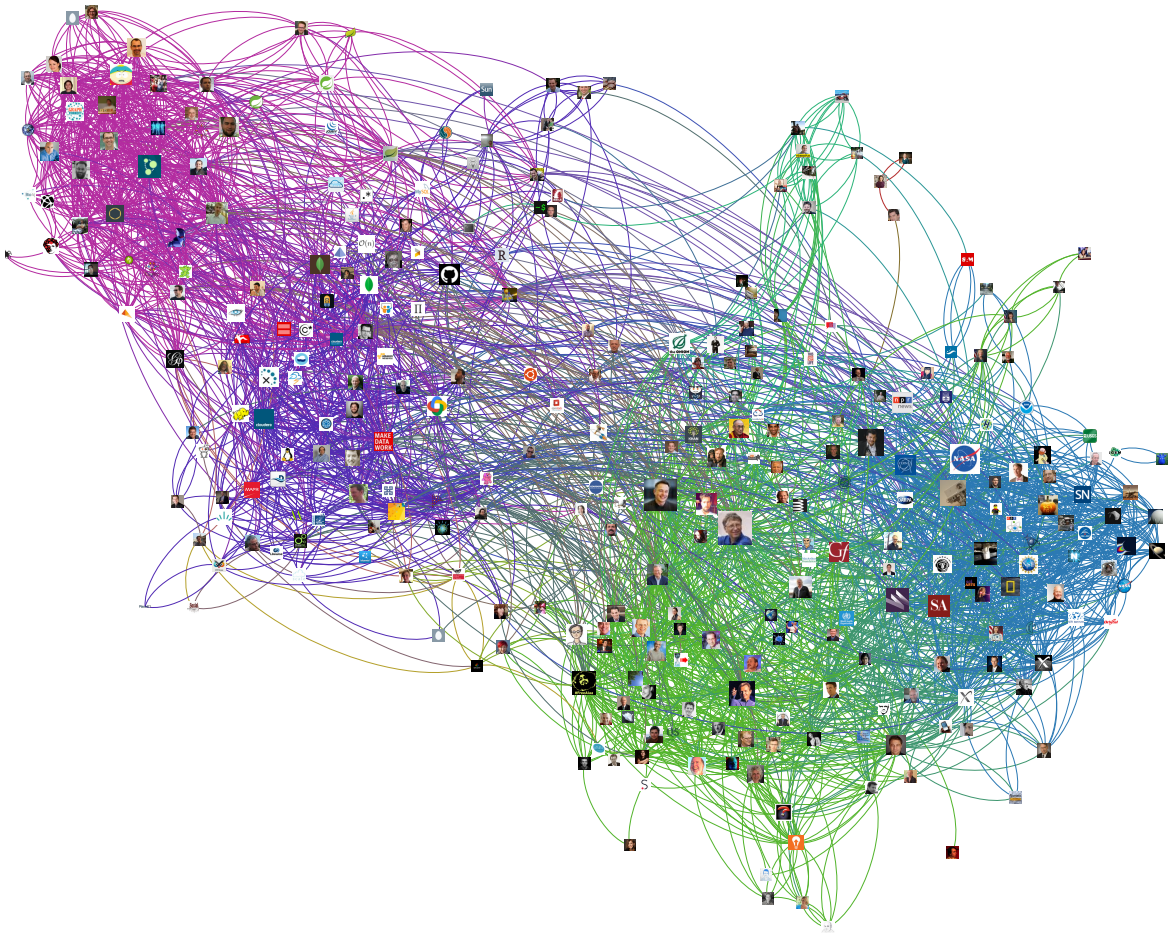
1. Introduction to network analysis
2. Longitudinal (dynamic) networks
3. The concept of new edges
4. Statistical anomaly-detection
5. Application to cyber-security
6. Some ideas for epidemiology applications

What is a network?

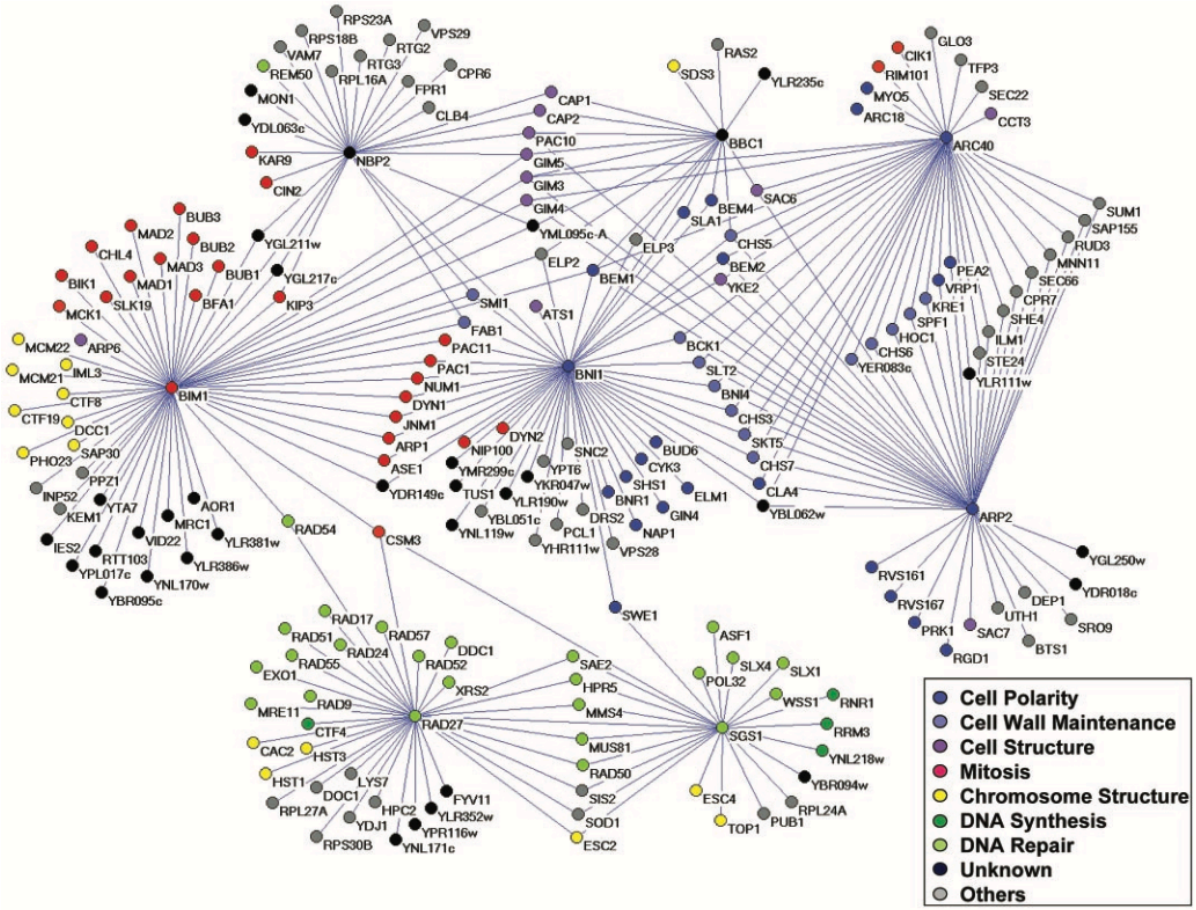


- ▷ a graph $G = (E, V)$: models the relationships between entities
- ▷ directed, undirected, weighted
- ▷ contributions from cell biology, engineering, epidemiology, neuro-science, sociology etc.

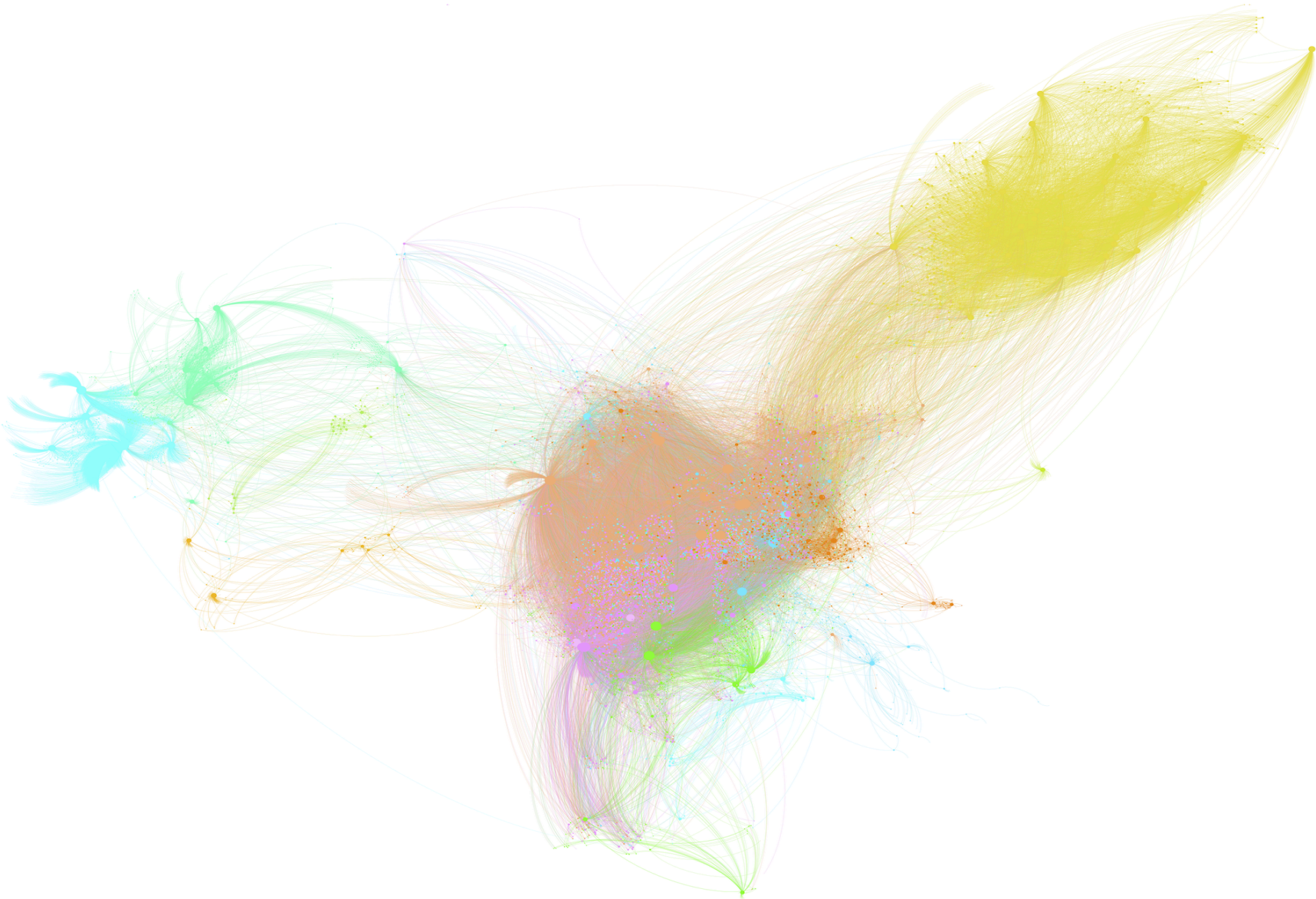
Examples: Social network (Twitter)



Examples: Genome network



Examples: Computer network (10min of traffic!)

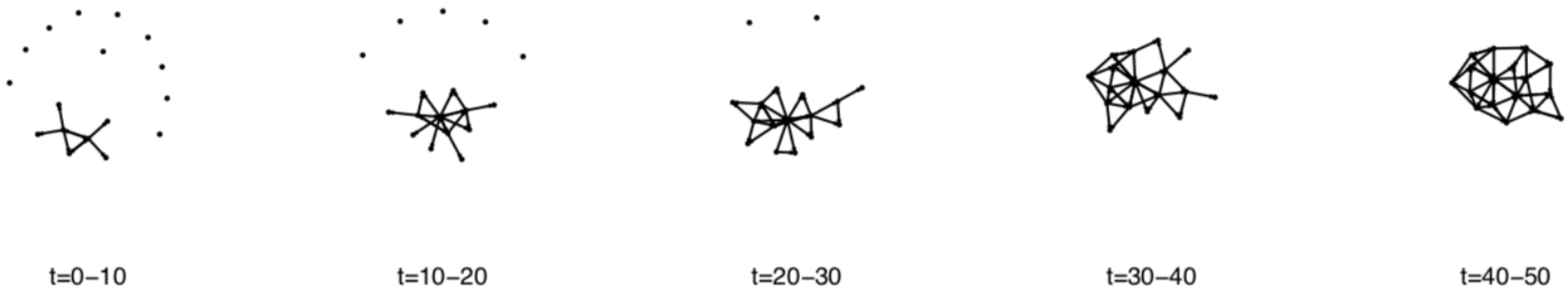


Static *vs.* dynamic

Static view: snapshot at a specific point in time ($\rightarrow$ cross-sectional data)

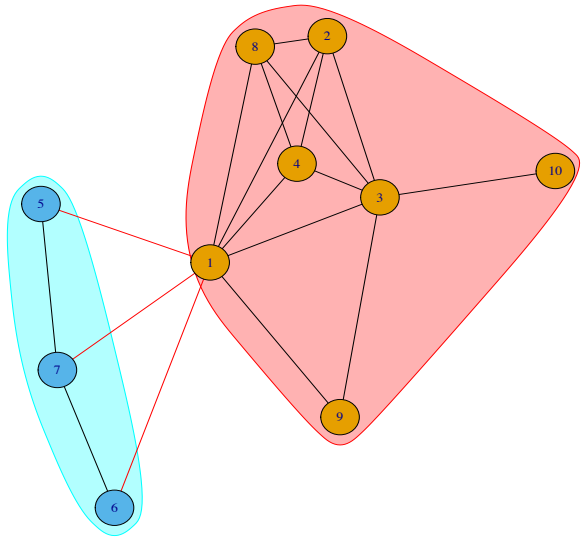
Dynamic view:

- follows the temporal evolution ($\rightarrow$ time-series data)
- fixed set of nodes
- evolving set of edges

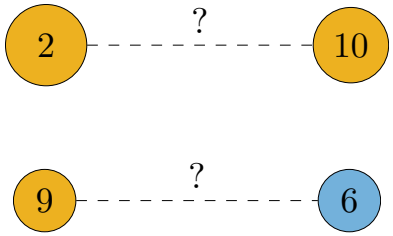


New edge prediction

New edges: edges not previously observed
Aim: predict new edges based on **existing structure** + **popularity**:
$\mathbb{P}\{\text{new edge at } t+1 | G_t\}$



How likely it is to observe the followings?

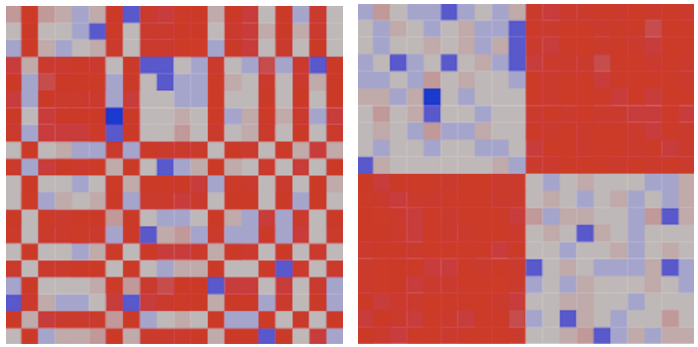


we propose a [Bayesian] PH model (Cox, 1972) for the conditional intensity, $\lambda_{ij}(t)$, of observing a new edge (i, j) after time t

Network structure (e.g. bi-clusters)

Adjacency matrix $A = \begin{pmatrix} 1 & 0 & \dots & 0 \\ 0 & 1 & 0 & \dots \\ \vdots & \vdots & \ddots & 0 \\ 1 & 0 & \dots & 1 \end{pmatrix}$

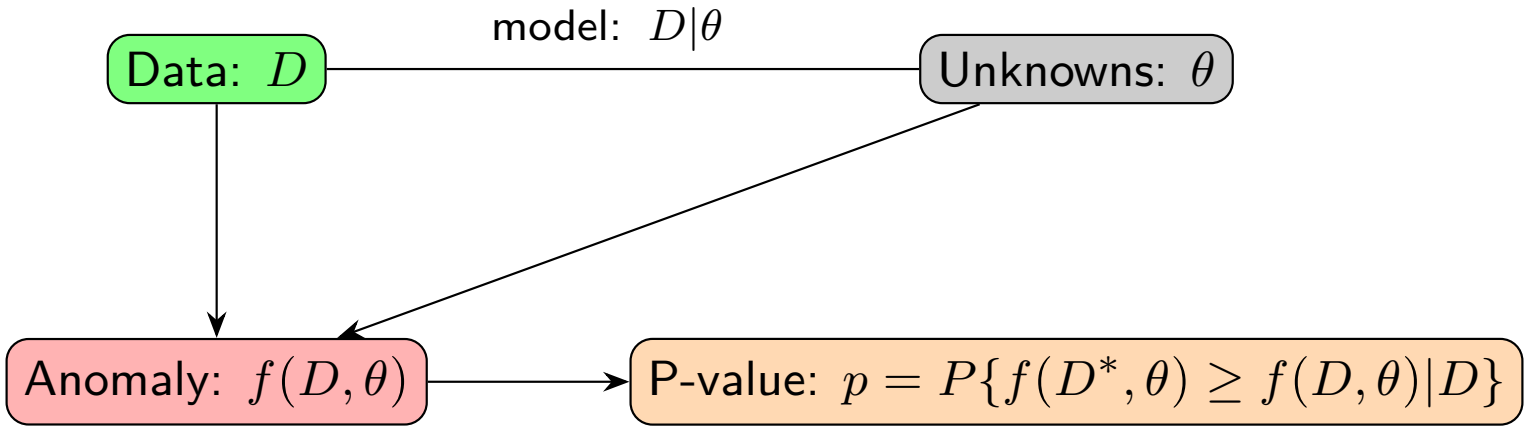
$$A(i,j) = \begin{cases} 1 & \text{if} \quad i \leftrightarrow j \\ 0 & \text{otherwise} \end{cases}$$



Original Reordered

- ▷ model-based hard-clustering (model-based similarity measures)
- ▷ heuristic dimensionality reduction (SVD, Spectral Clustering, K-means, PCA)
- ▷ flexible mixed-memberships/latent-features (LDA, Indian Buffet Process etc.)

Model-based anomaly-detection



$\mathbf{H}_0$: normal behaviour
$\mathbf{H}_1$: departure from model of normality learned

Anomaly: observation not compliant to learned network behaviour

Research can be performed from different perspectives:

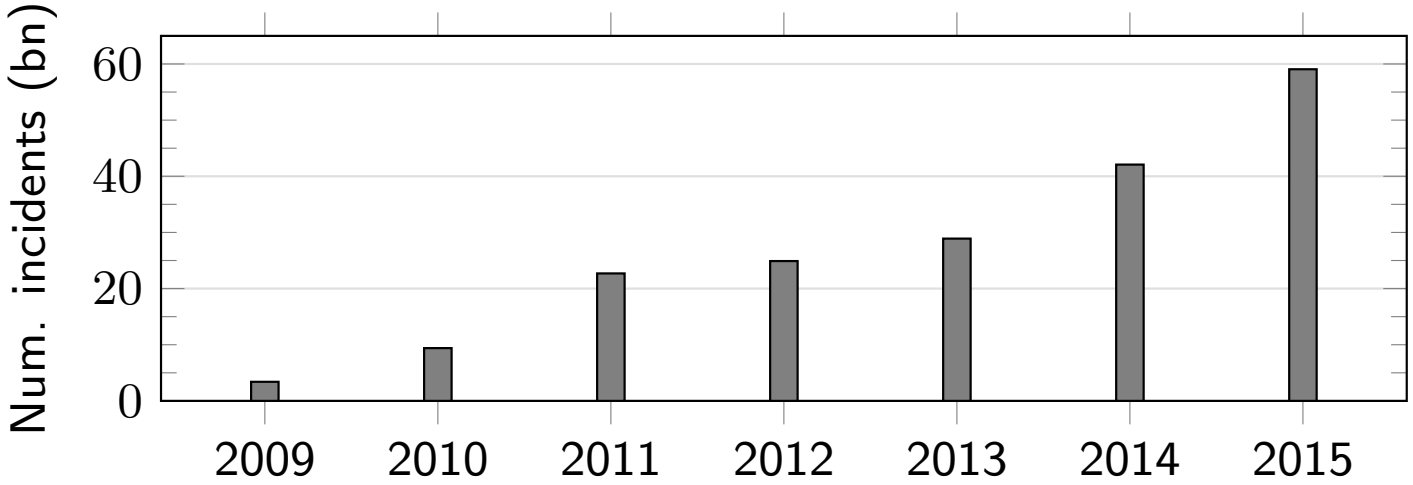
- whole graph (network structure)
- edges, nodes (temporal monitoring)

Cyber-security application

Context

Motivation: increasingly sophisticated, multi-stage cyber-attacks
e.g. *WannaCry 2017: 230,000 computers in 150 countries*

Global number of cyber incidents 2009-2015 (source: PwC)



(UK invested ∼1.9b in 2018)

Objective: monitor enterprise computer networks with really complex structure!

Challenges: computational speed and scalability (ideally, real-time detection)

Computer network data

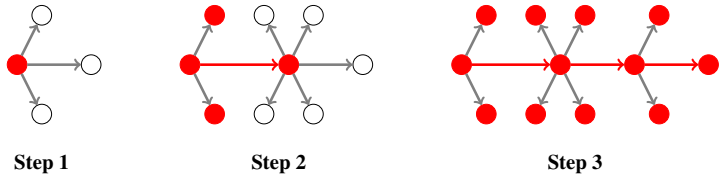
Variety:

- NetFlow data $\text{IP}_1 \rightarrow \text{IP}_2$ (time, duration, port, protocol, num. bytes...)
- Auth data (time, user, computer, success)
- Host-based data (run an agent on each host: time, process, parent process...)

Sources:

- Imperial College London (NetFlow)
- Los Alamos National Laboratory (90 days of traffic, 1 second resolution)

New edge intensity



New edges can be
- rarely, signal of anomaly
- regularly, formed by uninfected hosts

$$\lambda_{ij}(t) = \lambda(t)\exp\{\alpha \cdot (N_i^+(t), N_j^-(t), I_{i,1}(t), I_{i,2}(t)) + \beta_{ij} \cdot Z_{ij}(t)\} \times \mathbb{1}_{(I\times J)\setminus G_t\{(i,j)\}}$$

- $\lambda(t)$: 'seasonal' baseline (not parametrised)
- $N_i^+(t), N_j^-(t)$: time-varying in-degrees and out-degrees
- $I_{i,1}(t), I_{i,2}(t)$: time-varying indicators of new edge 'burstiness'
- $Z_{ij}(t)$: **matrix of attraction** $i \leftrightarrow j$ (similarity between nodes)

Cluster membership indicators

Dot-products of latent feature positions (IBP)

Application to Computer Network Data

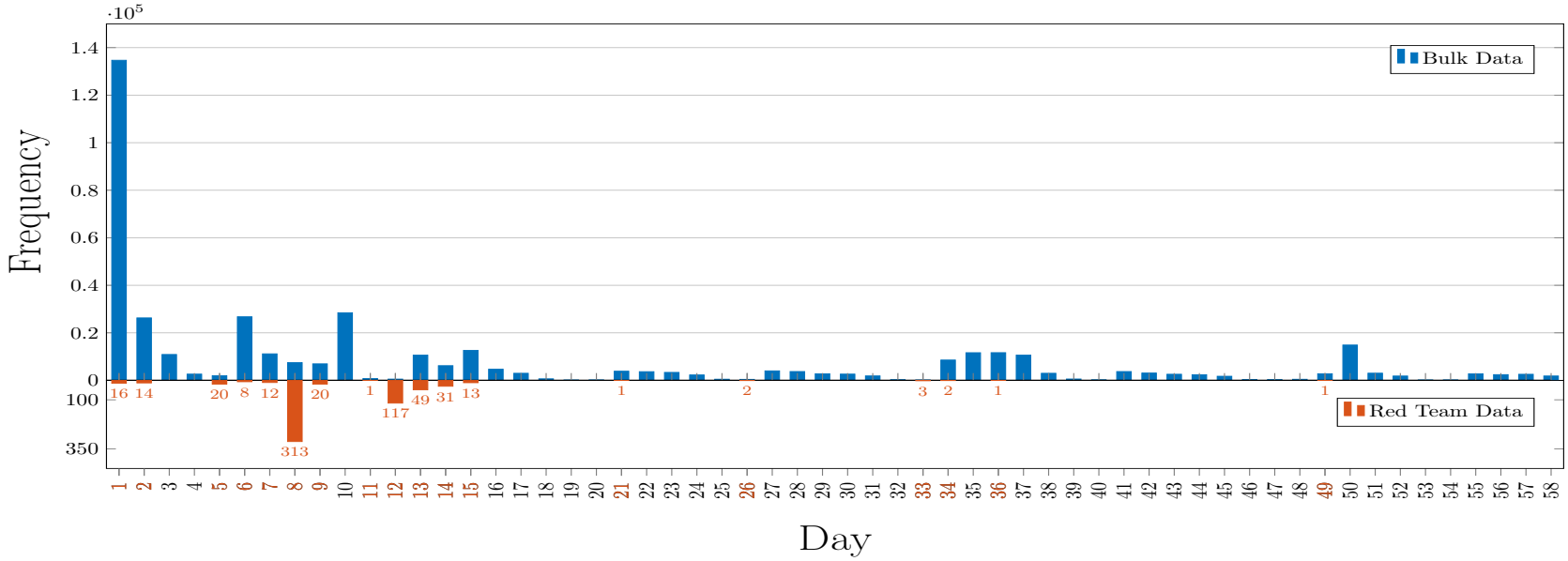
The LANL (Los Alamos National Laboratory) Data Set

Bulk:

- 1,648,275,307 events in total (58 days of traffic)
- 16,230 clients – 15,417 servers

Red Team:

- penetration testing: subset labelled as known compromised events
- 48,079 of the total records: 4 compromised clients



Model Prediction Performance

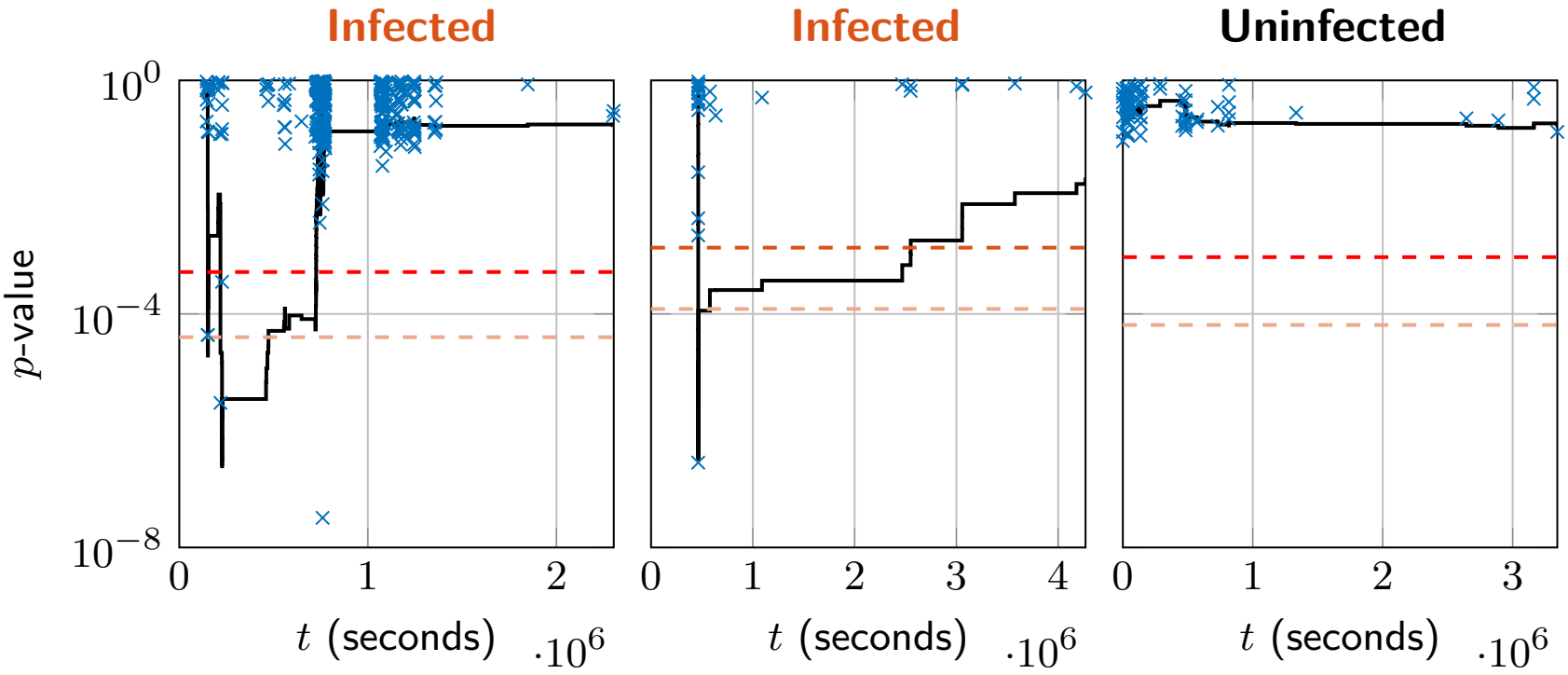
Method tested under both cluster and latent formulation:

- positive model coefficients: strong impact of degree and latent structure
- out-of-training log likelihood on the last 10,000 events

Model	**Log Likelihood**	**Iteration Time**
Cluster	-18804.34	81.4s
Latent-feature (IBP)	-18379.93	131.7s

We find that the latent feature model outperforms the cluster model

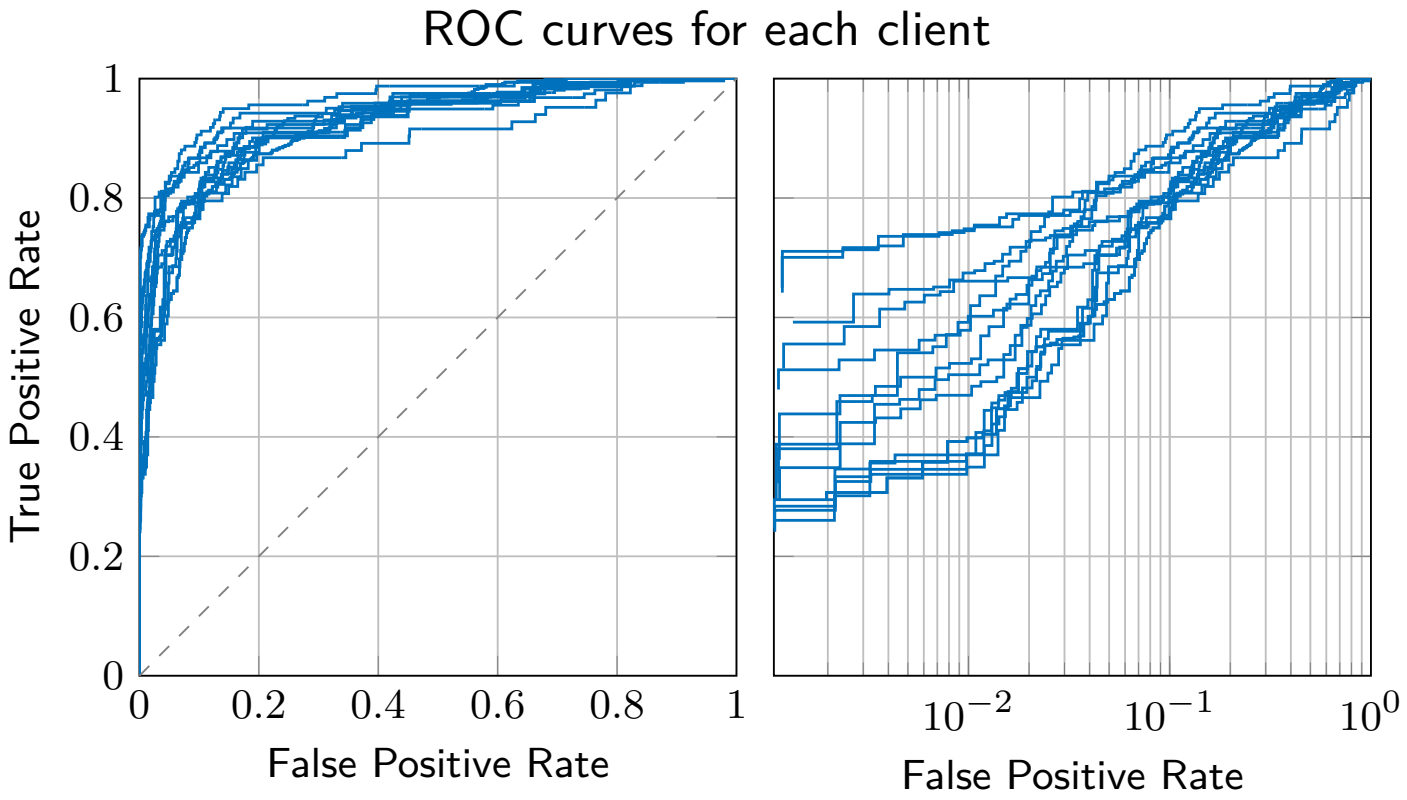
Anomaly-detection: control-charts



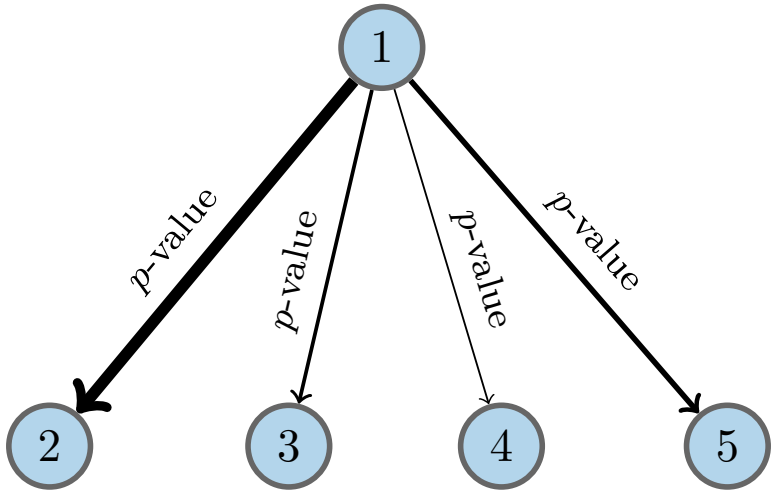
$$p_n = \frac{\sum_{(ij)\notin G_{t'_n}} \lambda_{ij}(t'_n)\mathbb{1}_{(0,\lambda_{i'_n j'_n}(t'_n)]}\{\lambda_{ij}(t'_n)\}}{\sum_{(i,j)\notin G_{t'_n}} \lambda_{ij}(t'_n)}$$

$$s_i(t) = \bar{\chi}^2_{2\{1+N_i^+(t)\}}\left(-2\sum_{n\geq 1}\mathbb{1}_{[0,t)}(t'^i_n)\log p^i_n\right) \text{ s.t. } \inf_{t\geq 0} s_i(t)$$

Anomaly-detection: ROC curves



Challenges

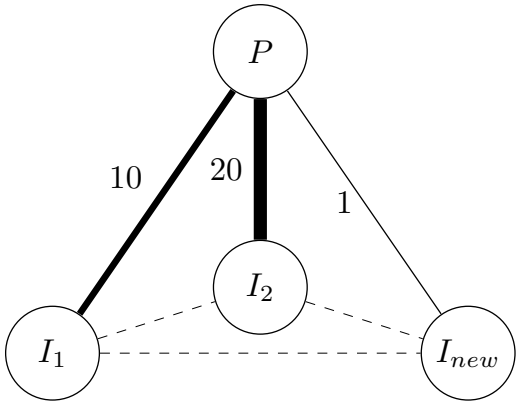


1. Robust model for predicting links/new links
2. How to robustly combine p-values (avoiding conservative tests)?
3. How to take into account different node degrees/edge weights?

How to exploit all of this in medical research?

Static prediction of clinical interventions

Suppose we have a network of interventions for a certain disease:



- compare interventions (cluster 'similar' interventions)
- Bayesian approach is attractive:
  - if meta-analysis is based only on few studies
  - if new study estimates different functions $\rightarrow$ combine multiple into single param. (hierarchical random-effects models)
  - only option when very complex models are used

How to exploit all of this in medical research?

Real-time prediction of clinical interventions

Monitor networks of interventions in ICUs (noisy, sparse, heterogeneous data)

- predict onset of multiple invasive treatments
- other similar survival analysis/ Cox models
    - static covariates: gender, age etc..
    - time-varying covariates: oxygen saturation, blood urea nitrogen etc..

Monitor networks of personalised healthcare

- high-frequency data (from mobile-devices etc..)
- descriptions of symptoms over time (text-mining)

Final Remarks

- Introduction to network analysis (static/dynamic)
- Prediction of new edges in a large network (based on network structure)
- Provide some insights on statistical anomaly-detection
- Results from cyber data $\rightarrow$ show good performance
- Some opportunities in the field of medical research
- Challenges still to be addressed:
  - effectively combine information from different edges (interventions)
  - Bayesian inference (MCMC): need fast methods

References

- Cox, D. R. (1972). "Regression models and life-tables (with discussion).". In: *Journal of the Royal Statistical Society, Series B* 34, pp. 187–220.
- Heard, N. and S. Metelli (2014). "Modelling new edge formation in a computer network through Bayesian variable selection". In: *Joint Intelligence and Security Informatics Conference (JISIC), 2014 European*. IEEE, pp. 272–275.
- Heard, N and P Rubin-Delanchy (2018). "Choosing between methods of combining p-values". In: *Biometrika* 105, pp. 239–246. DOI: biomet/asx076.
- Heard, N, P Rubin-Delanchy, and D Lawson (2014). "Filtering automated polling traffic in computer network flow data". In: IEEE, pp. 268–271. DOI: 10.1109/JISIC.2014.52.
- Metelli, S and NA Heard (2016). "Model-based clustering and new edge modelling in large computer networks". In: IEEE. DOI: 10.1109/ISI.2016.7745449.
- Metelli, S. and N.A. Heard (2018). "On Bayesian new edge modelling and anomaly detection in computer networks". Submitted to the Annals of Applied Statistics.

Thank you!