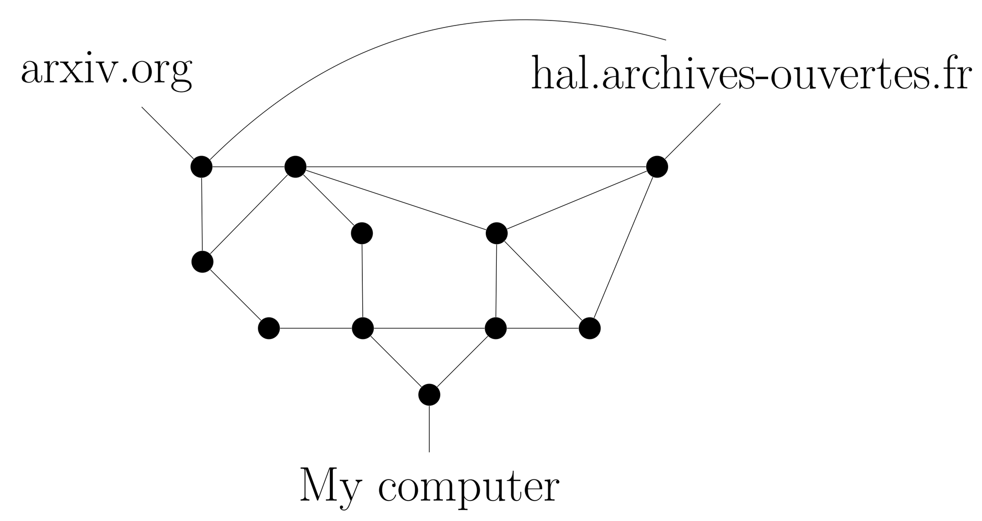
Internet Topology Dynamics:
stochastic process estimation from partial observations

Sergey Kirgizov, Clémence Magnien

LIP6 (CNRS – UPMC) Paris

Internet topology



nodes : computers
links : connections between computers

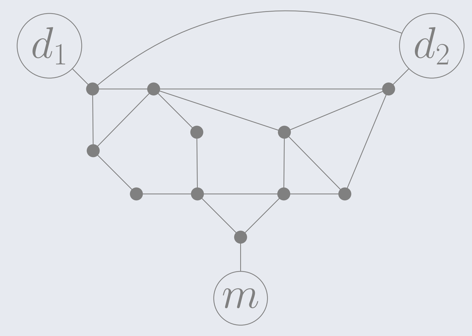
Static case

Long and expensive measures
Bias in the observed structure $\Big\} \Longrightarrow$ No reliable map

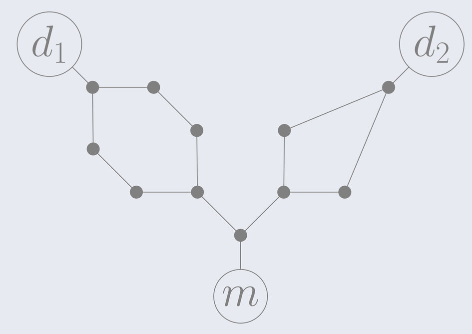
Dynamics

All static problems
Topological changes $\Big\} \Longrightarrow$ Not so easy

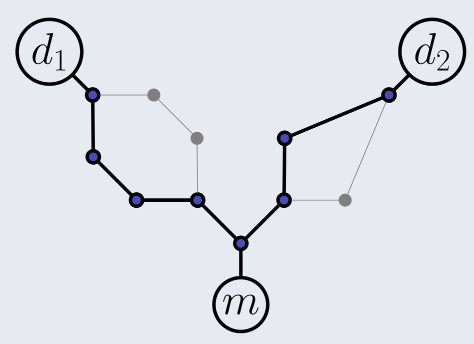
Egocentric view



Real topology

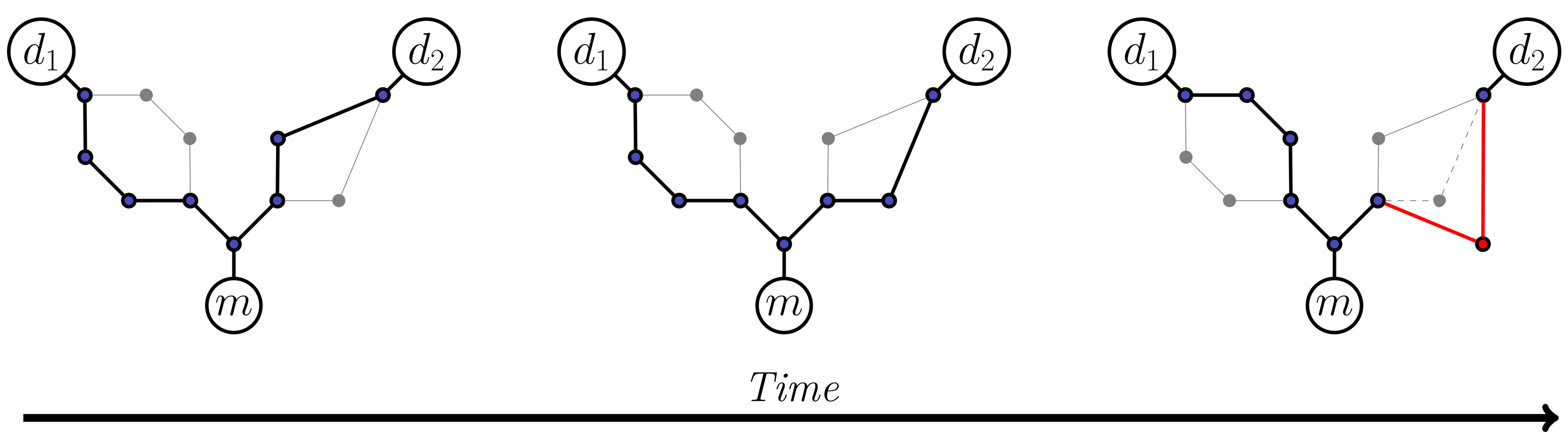


(Shortest) routes between monitor and destinations



One measurement by `tracetree`

Dynamics of egocentric views



Repeated measurements $\Longrightarrow$ study of dynamics

Stochastic inference from partial observations

X – stochastic process
$\langle X\rangle$ – "realization" of X
$\langle Y\rangle$ – observations, measurements

$\langle Y\rangle \subset \langle X\rangle$

$$\begin{tikzcd} X \arrow[r, "r"] & \langle X\rangle \arrow[d, "o"] \\ & \langle Y\rangle \arrow[ul, dashed, "?"] \end{tikzcd}$$

Note that a "direct" map $\langle Y\rangle \to \langle X\rangle$ does not exist. We absolutely have to guess X firstly.

1-point process. Poissonian case

We have one object. The object can change at some moment. The object never changes back. We suppose that number of changes that occurs during Δ follows some poissonian law (parameterized by λ, i.e. mean number of changes in Δ). How can we infer the most likely λ having only a sequence of observations?

$$\begin{array}{ll} \text{Real } \langle X\rangle: & \longrightarrow \bullet' \longrightarrow \bullet'' \longrightarrow \bullet''' \longrightarrow \cdots \\ \text{Observed } \langle Y\rangle: & \xrightarrow{\textit{change}} o' \xrightarrow[\textit{nothing}]{} o' \xrightarrow{\textit{change}} o'' \longrightarrow \cdots \end{array}$$

N – number of intervals
c – number of intervals with at least one change

$1 - \mathrm{Pois}_0(\lambda)$ – the probability of "at least one change"

$$1 - \mathrm{Pois}_0(\lambda) \approx \frac{c}{N}$$

$$\hat{\lambda} = \mathrm{Pois}_0^{-1}\left(1 - \frac{c}{N}\right)$$

Internet dynamics is not 1-point poissonian process

This naïve 1-point poissonian approach does not work with internet dynamics:

! Due to load-balancing old states can reoccur in $\langle Y\rangle$.
! The internet topology is not a one solid object.
! Different parts of topology have different dynamics.

n-point process

Contrary to 1-point process, n-point process deals with a set U of points. At each time we observe only a random part of U.

$$\begin{array}{ll} \text{Real } \langle X\rangle: & (a,b,c) \longrightarrow (a,b,d,e) \longrightarrow (e,b,c) \longrightarrow (e,b,d) \longrightarrow (e,b,d) \\ \text{Observed } \langle Y\rangle: & (a,b) \longrightarrow (e,c) \longrightarrow (e,b) \end{array}$$

Different parts have different dynamics

$\bullet$ – part of a topology (e.g. a node)
δ – lifetime of $\bullet$
p – probability of being observed
Δ – interval between observations

Denote by o the number of observations that contains our $\bullet$. Using our sequence of observations, we approximate:

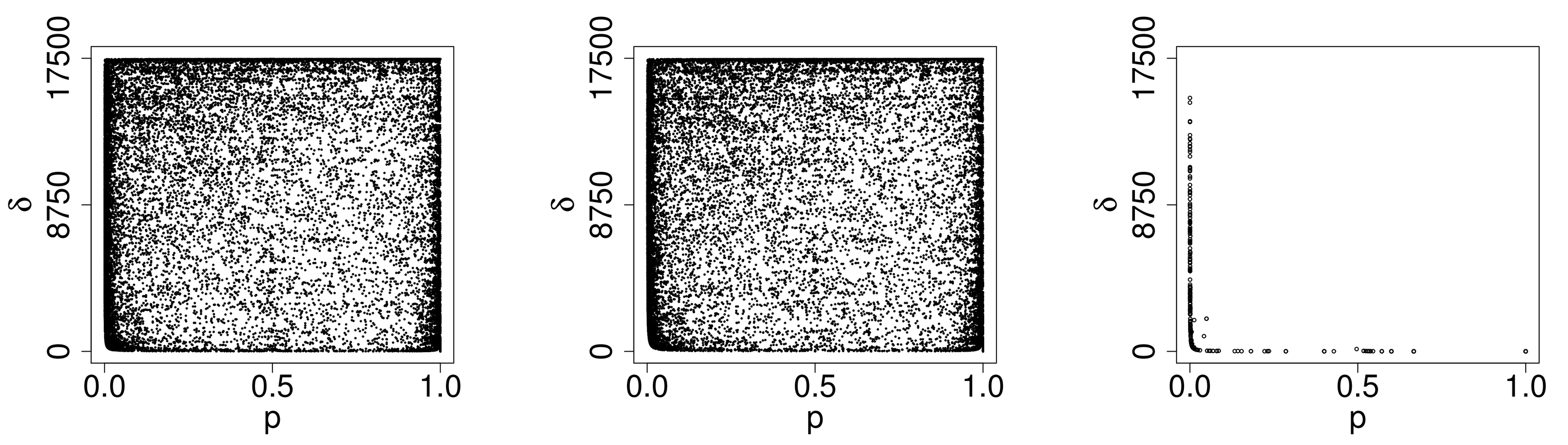
$$\delta = t_{\text{last}} - t_{\text{first}}, \quad p = \frac{o}{n}.$$

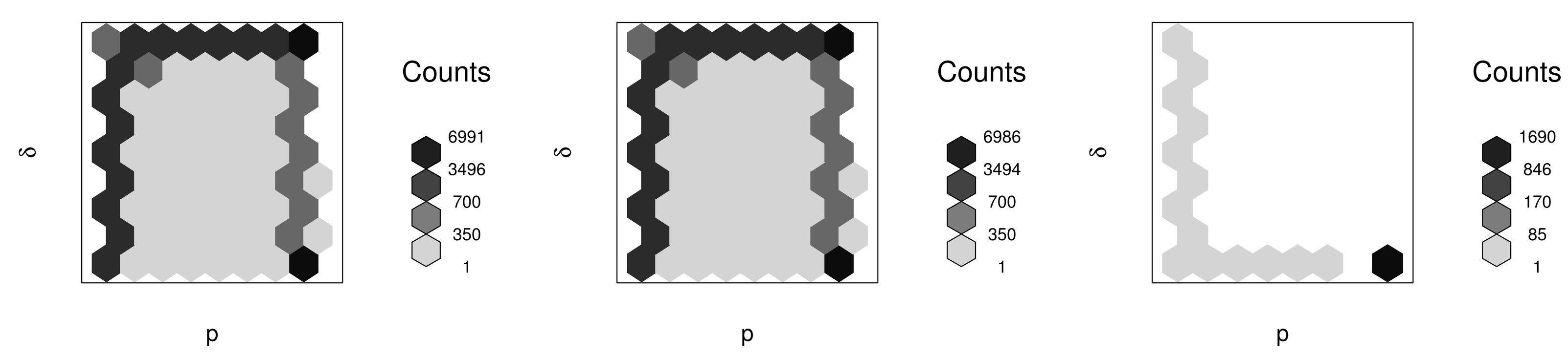
During the lifetime of $\bullet$ we perform $n = \frac{\delta}{\Delta}$ observations. Suppose that p is constant over the lifetime of $\bullet$. Now we write the probability that particular $\bullet$ is missed.

$$\Pr_{\text{mis}} = (1-p)^n.$$

From this we can conclude that the missed parts of topology have a short lifetime or have a very small chance of being observed.

$\langle X\rangle \xrightarrow{\Delta} \langle Y\rangle_1$ | $\langle X\rangle \xrightarrow{2\Delta} \langle Y\rangle_2$ | $\langle Y\rangle_1 - \langle Y\rangle_2$





Towards a solution of n-point process

Possible approach consists in constructing a transformation:

$$n\text{-point process} \overset{?}{\dashrightarrow} 1\text{-point process}$$

But it is not easy, particularly when points have different properties, e.g. different probability of being observed.

Conclusion

- Internet topology dynamics can be modelled as partially observed n-point stochastic process (where n is not a constant).
- Different parts of the internet have different dynamics.
- Using our measurements we miss only the nodes with short lifetime or with very small probability of being observed.

Open questions:

- Trasformation "n-point process $\overset{?}{\dashrightarrow}$ 1-point process".
- Good canditate for X (Hawkes processes?).
- Universal inference $\langle Y\rangle \overset{?}{\dashrightarrow} X$.