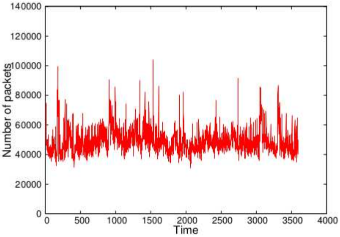
Stream graphs for modelling temporal networks

Tiphaine Viard

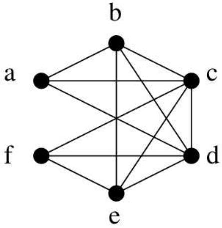
Peyresq 2023

Context

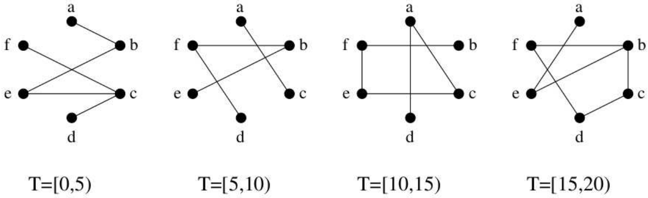
Time (signal processing) :



Structure (graph theory):



"Time" + structure (sequence of graphs)



Goal

Assessment : progress **limited** by **fundamental** locks

→ Loss of information, inadequate formalism...

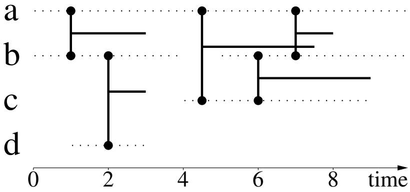
Our goal :
A language for interactions
comparable to graph theory
for networks

- **Simple** and **intuitive**
- Generalizes **graphs** and **signal**
  - degree? clustering? autocorrelation? Fourier transform? ...?
- Allows **applicative progress**

Stream graphs

$S = (T, V, W, E)$

$T = [\alpha, \omega]$ $\quad V = \{u\}$ $\quad W = \{(t, v)\}$ $\quad E = \{(t, uv)\}$



c is present from time 4 to time 9: $\{c\} \times [4, 9] \in W$

a interacts with b from time 1 to time 3: $\{ab\} \times [1, 3] \in E$

...

T_u = presence of node u

$\rightarrow T_b = [0, 4] \cup [4.5, 10]$

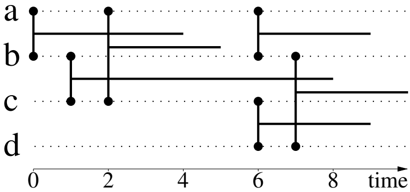
T_{uv} = presence of link uv

$\rightarrow T_{ab} = [1, 3] \cup [7, 8]$

Stream graphs

$S = (T, V, W, E)$

$T = [\alpha, \omega]$ $V = \{u\}$ $W = \{(t, v)\}$ $E = \{(t, uv)\}$



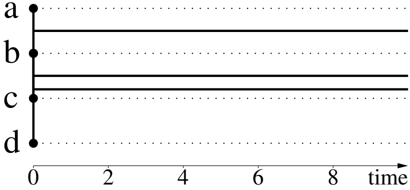
c is present from time 4 to time 9: $\{c\} \times [4, 9] \in W$
a interacts with b from time 1 to time 3: $\{ab\} \times [1, 3] \in E$
...

T_u = presence of node u
$\rightarrow T_b = [0, 4] \cup [4.5, 10]$
T_{uv} = presence of link uv
$\rightarrow T_{ab} = [1, 3] \cup [7, 8]$
If $\forall v$, $T_v = T$, then **link stream**

Stream graphs

$S = (T, V, W, E)$

$T = [\alpha, \omega]$ $V = \{u\}$ $W = \{(t, v)\}$ $E = \{(t, uv)\}$



c is present from time 4 to time 9: $\{c\} \times [4, 9] \in W$
a interacts with b from time 1 to time 3: $\{ab\} \times [1, 3] \in E$
...

T_u = presence of node u
$\rightarrow T_b = [0, 4] \cup [4.5, 10]$
T_{uv} = presence of link uv
$\rightarrow T_{ab} = [1, 3] \cup [7, 8]$
If $\forall v, T_v = T$ and $\forall u, v, T_{uv} = \{\emptyset, T\}$, then **graph-equivalent**

Substreams and clusters

Graphs:

Subgraph $= G' = (V', E')$ such that $V' \subseteq V, E' \subseteq E$

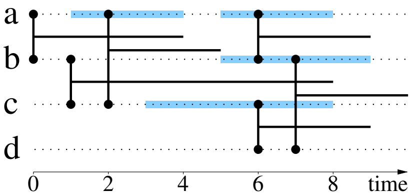
Cluster of nodes $C =$ set of nodes

Stream graphs:

Substream $= S' = (T', V', W', E')$ such that

$T' \subseteq T, V' \subseteq V, W' \subseteq W, E' \subseteq E$

Cluster of nodes $C =$ set of (t, v)



Density

Graphs :

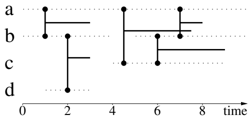
u, v random = link?

$$\delta(G) = \frac{2m}{n \cdot (n-1)}$$

Stream graphs:

uv, t random = link?

$$\delta(S) = \frac{\sum_{uv \in V \otimes V} T_{uv}}{\sum_{uv \in V \otimes V} T_u \cap T_v}$$



Graph-equivalent streams: $\delta(S) = \delta(G)$

Density

Graphs :

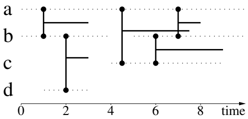
u, v random = link?

$\delta(G) = \frac{2m}{n \cdot (n-1)}$

Stream graphs:

uv, t random = link?

$$\delta(S) = \frac{\sum_{uv \in V \otimes V} T_{uv}}{\sum_{uv \in V \otimes V} T_u \cap T_v}$$



$$\delta(S) = \frac{10}{22} \approx 0.45$$

Graph-equivalent streams: $\delta(S) = \delta(G)$

Neighborhood

Graphs :

Neighborhood = set of nodes

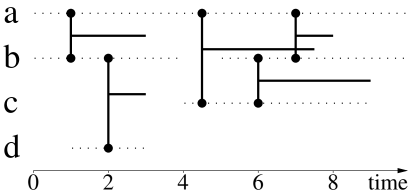
$N(u) = \{v : uv \in E\}$ $d(u) = |N(u)|$

Stream graphs:

Neighborhood = cluster

$N(u) = \{(t, v) : (t, uv) \in E\}$

$d(u) = \frac{|N(u)|}{|T|}$



Graphs, stream graphs: $\sum_u d(u) = 2m$

Neighborhood

Graphs :

Neighborhood = set of nodes

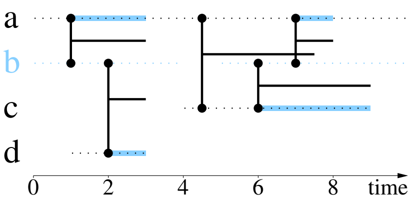
$N(u) = \{v : uv \in E\}$ $d(u) = |N(u)|$

Stream graphs:

Neighborhood = cluster

$N(u) = \{(t, v) : (t, uv) \in E\}$

$d(u) = \frac{|N(u)|}{|T|}$



Graphs, stream graphs: $\sum_u d(u) = 2m$

Degrees

$$d(v) = \frac{|N(v)|}{|T|} = \int_t \frac{d_t(v)}{|T|} dt$$

$$d(t) = \sum_v \frac{d_t(v)}{|V|}$$

$$\hat{d}(v) = \int_t \frac{d_t(v)}{|T_v|} dt$$

$$\hat{d}(t) = \sum_v \frac{d_t(v)}{|V_t|}$$

$$d(V) = \sum_{v \in V} \frac{|T_v|}{|W|} d(v)$$

$$d(T) = \int_t \frac{|V_t|}{|W|} d(t) dt$$

$$d(S) = \sum_v \frac{1}{|V|} d(v) = \frac{2 \cdot |E|}{|T \times V|} = \int_t \frac{1}{|T|} d(t) dt$$

$$\hat{d}(S) = \frac{\sum_v \int_t d_t(v) dt}{|W|} = \frac{2 \cdot |E|}{|W|} = \frac{2m}{n}$$

Link stream

$\implies d(v) = \hat{d}(v), d(t) = \hat{d}(t), d(V) = d(T) = d(S) = \hat{d}(S)$

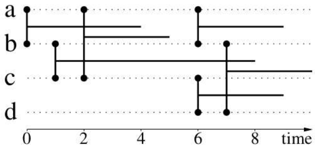
Clustering coefficient

Graphs:

$$cc(u) = \delta(N(u))$$

Stream graphs:

$$cc(u) = \delta(N(u))$$



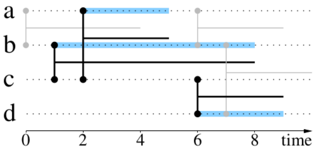
Clustering coefficient

Graphs:

$$cc(u) = \delta(N(u))$$

Stream graphs:

$$cc(u) = \delta(N(u))$$



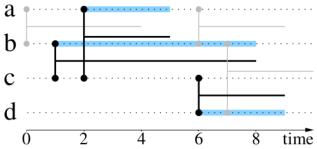
Clustering coefficient

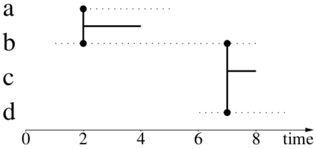
Graphs:

$$cc(u) = \delta(N(u))$$

Stream graphs:

$$cc(u) = \delta(N(u))$$





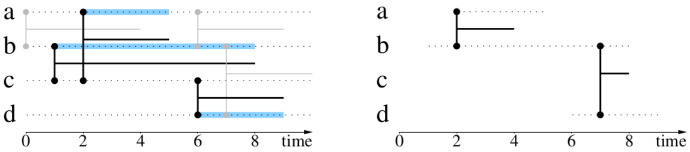
Clustering coefficient

Graphs:

$$cc(u) = \delta(N(u))$$

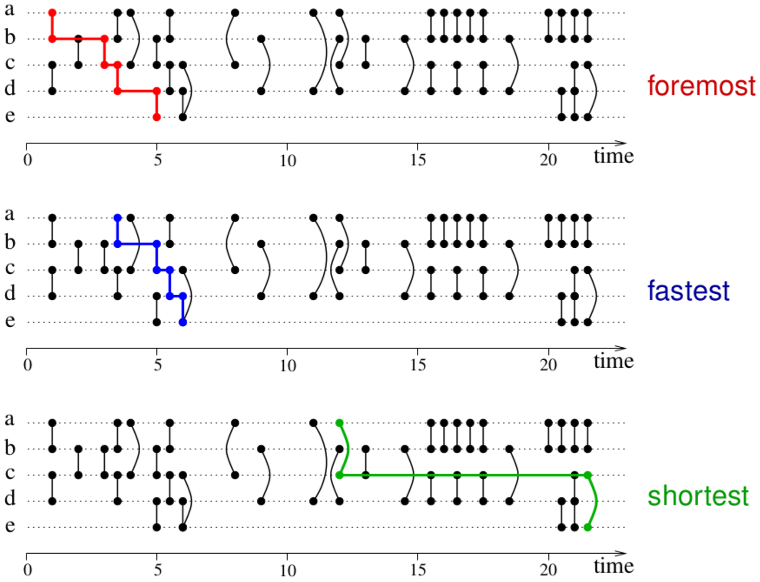
Stream graphs:

$$cc(u) = \delta(N(u))$$

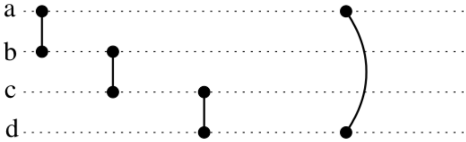


$$cc(c) = \frac{3}{5} = 0.6$$

Paths



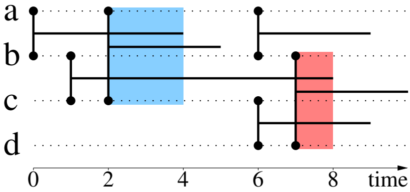
Monsters: connected parts



Cliques

Graphs:
Clique X = subgraph $G(X)$ with density 1
Max. if included in no other clique

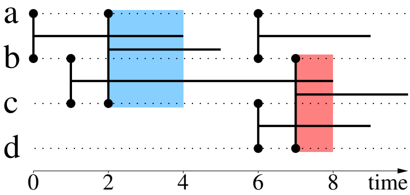
Stream graphs:
Clique X = cluster with density 1
Max. if included in no other clique



Cliques

Graphs:
Clique X = subgraph $G(X)$ with density 1
Max. if included in no other clique

Stream graphs:
Clique X = cluster with density 1
Max. if included in no other clique



Computing maximal compact cliques of S in $\mathcal{O}(2^n n^2 m^3 + 2^n n^3 m^2)$ time, and $\mathcal{O}(2^n n m^2)$ space

Defined notions

Fundamental notions

Stream graphs, link streams
Size, duration, uniformity and compacity
Substreams clusters

Density-based

Density, cliques
Neighbourhood, degree
Clustering, transitivity
Cluster relations, quotient stream
Line streams
k-cores

Path-based

Paths and distances
Connectivity
Connected components
Trees
Cascades
Centralities

Generalizations

Δ-analysis and instantaneous links
Bipartite, weighted, directed, multilayer streams

Clique enumeration algorithm

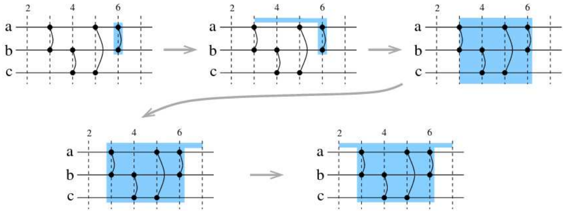
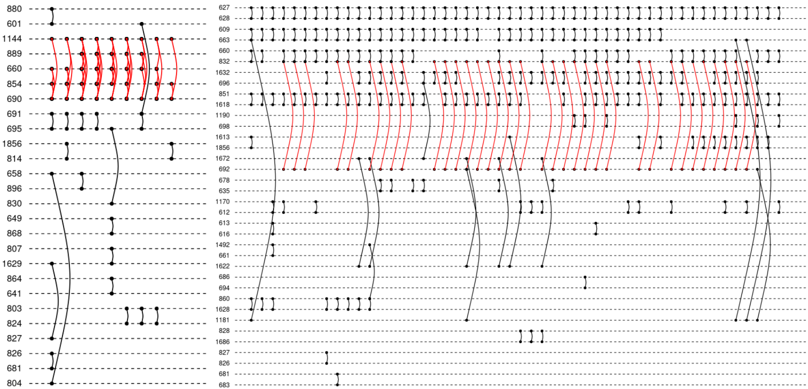


Fig. 2: A sequence of Δ-cliques built by our algorithm to find a maximal Δ-clique (bottom row) from an initial trivial Δ-clique (top-left) in the link stream of Figure 1 when $\Delta = 3$. From left to right and top to bottom: the algorithm starts with $(\{a, b\}, [6, 6])$, and finds $(\{a, b\}, [3, 6])$ thanks to lines 9 to 12 of the algorithm. It then finds $(\{a, b, c\}, [3, 6])$ thanks to lines 6 to 8. It finds $(\{a, b, c\}, [3, 7])$ from lines 13 to 16, and finally $(\{a, b, c\}, [2, 7])$ from lines 9 to 12.

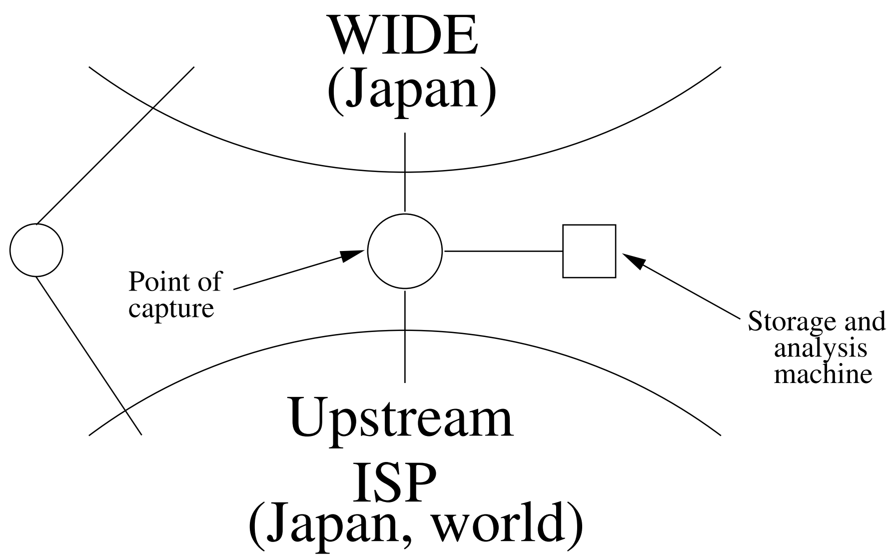
Some mobility results



Dataset

MAWI: 15 years of IP traffic captures

→ bipartite by measure



Strong need of progress

High volume of data → 88.10^6 links / hour

Cliques for anomaly detection

Clique = set of machines all interacting on a short period of time

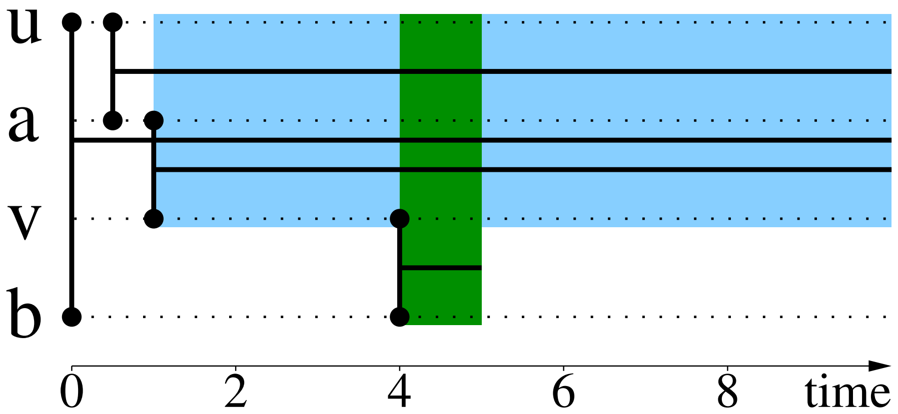
→ DDoS, scan, load balancing...

Enumeration out of reach

→ what makes a good sampling?

A **star** is a trivial clique

→ large balanced cliques

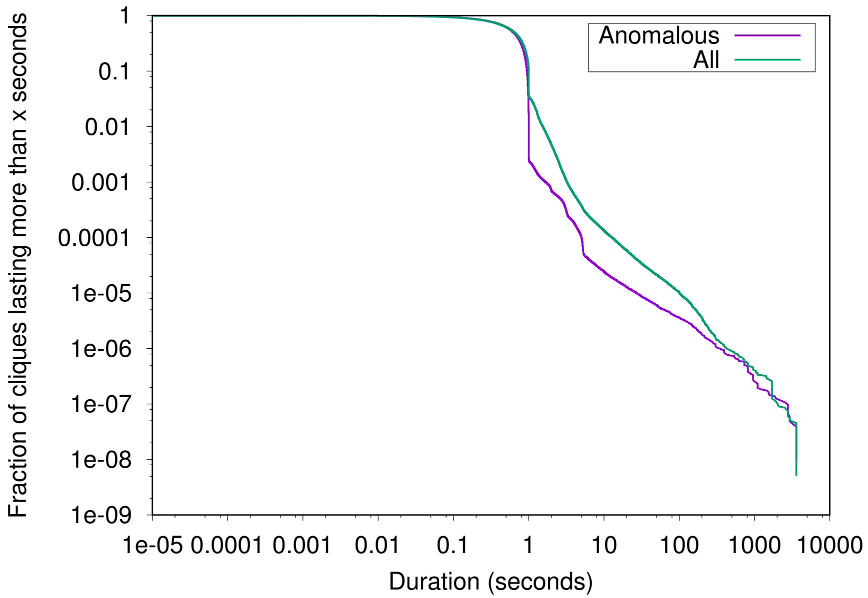


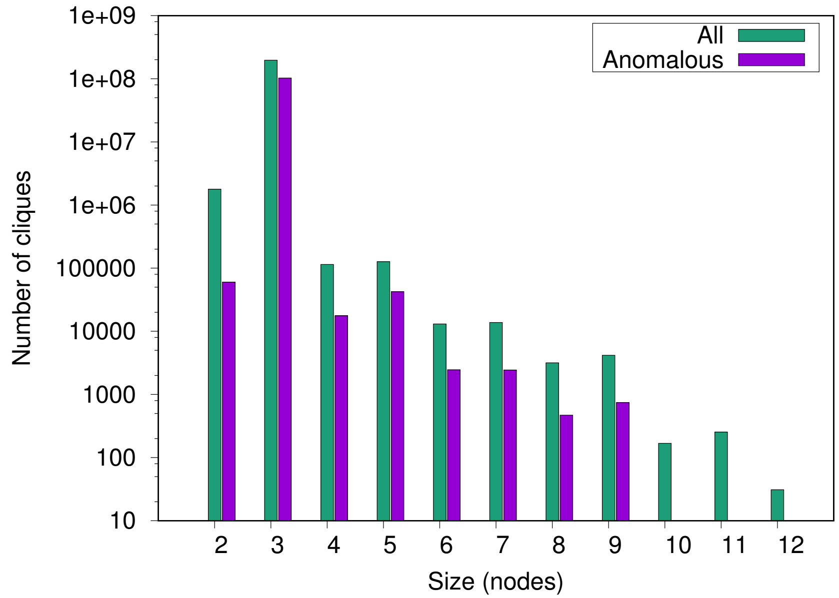
Sampled cliques

$1.3 \cdot 10^7$ cliques

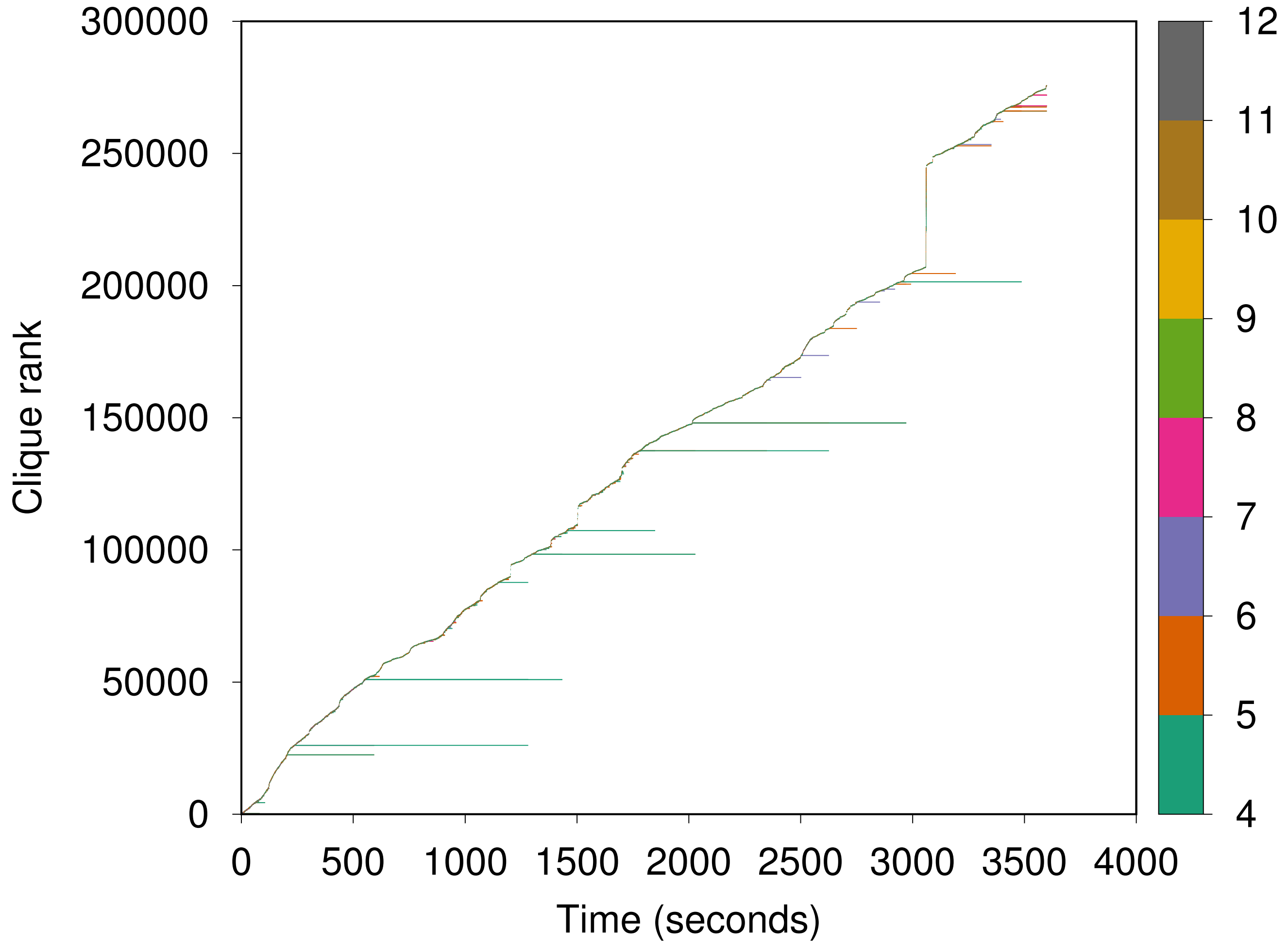
$1.9 \cdot 10^6$ distinct

106 days

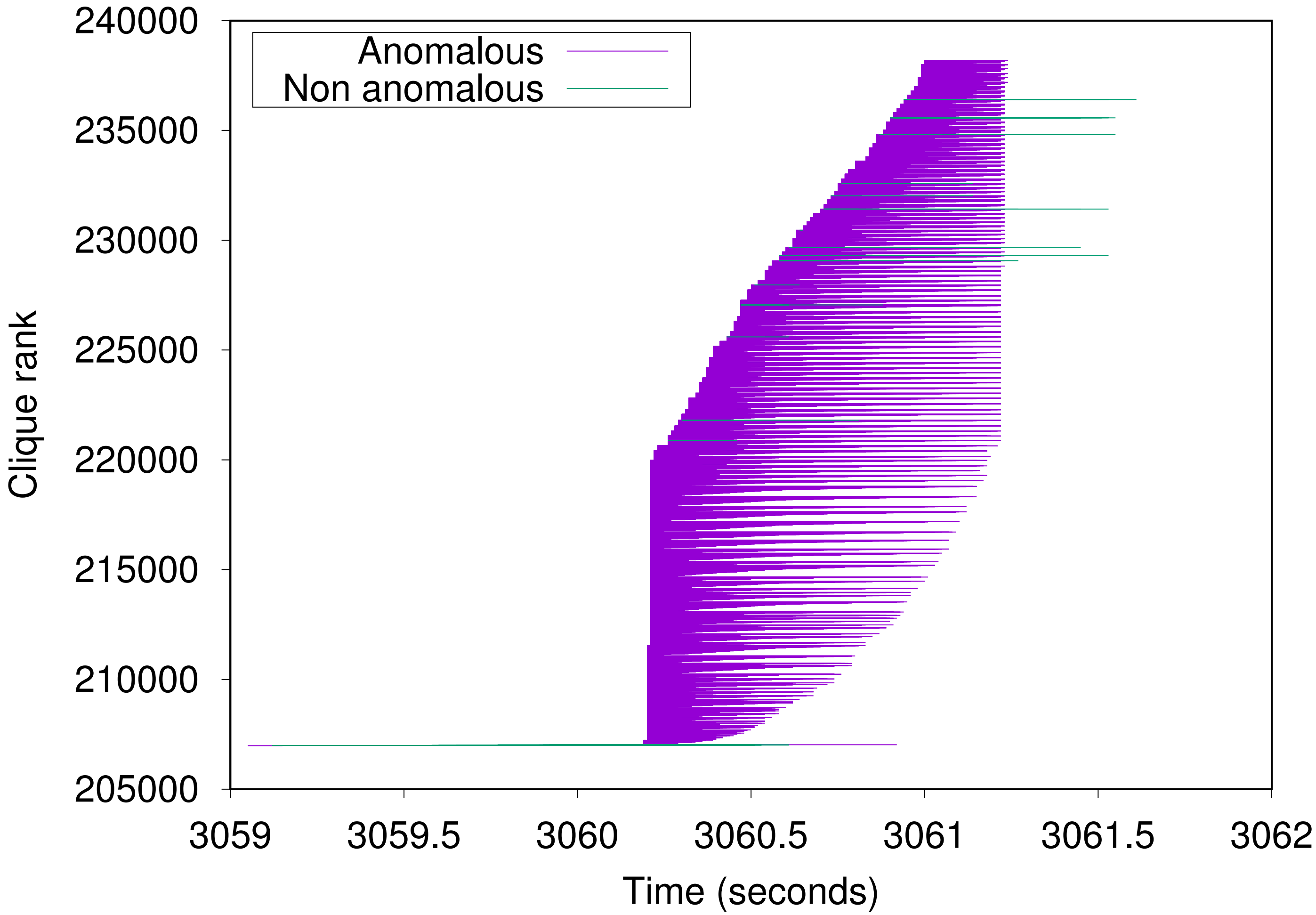




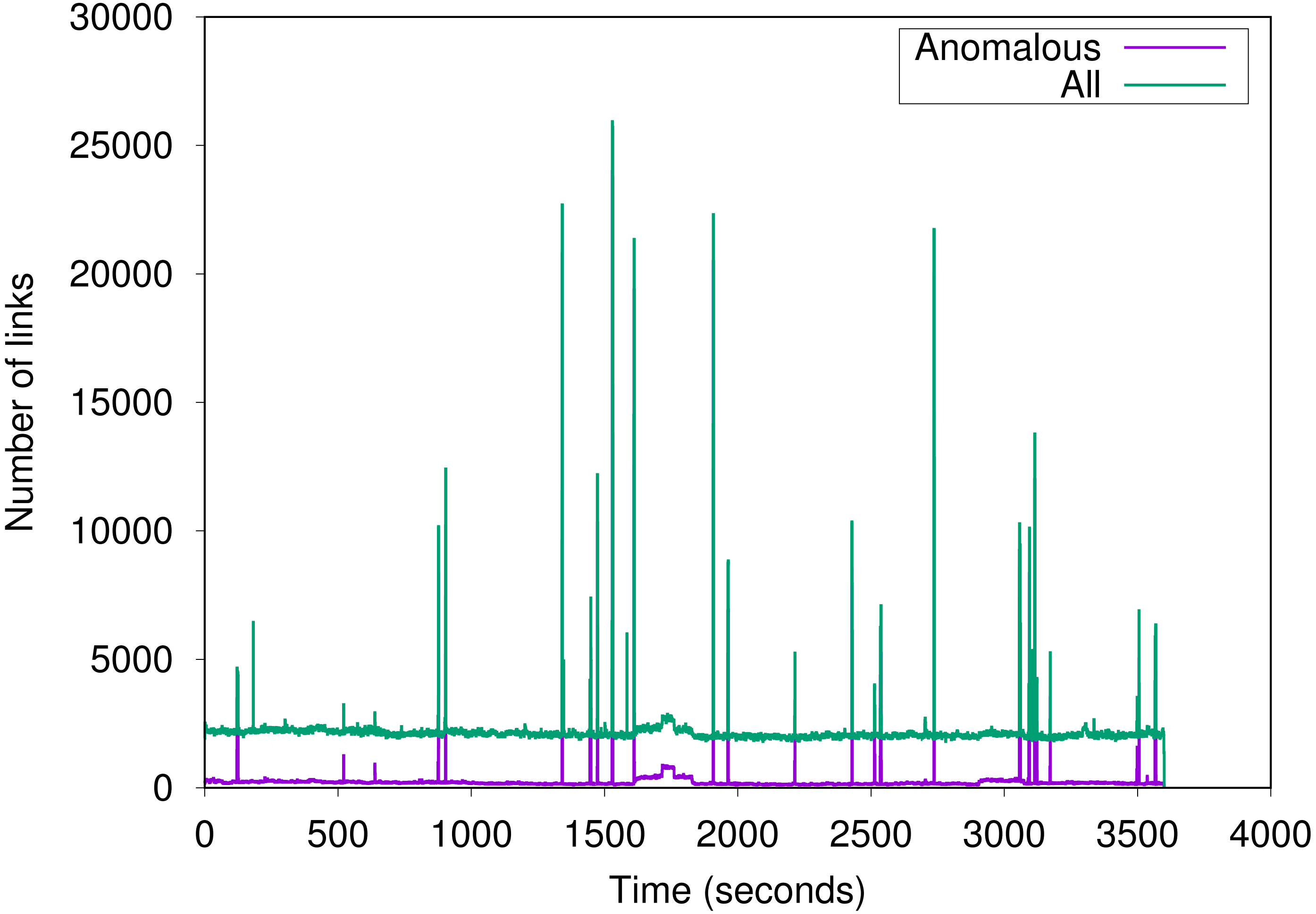
Cliques involving 4 nodes or more



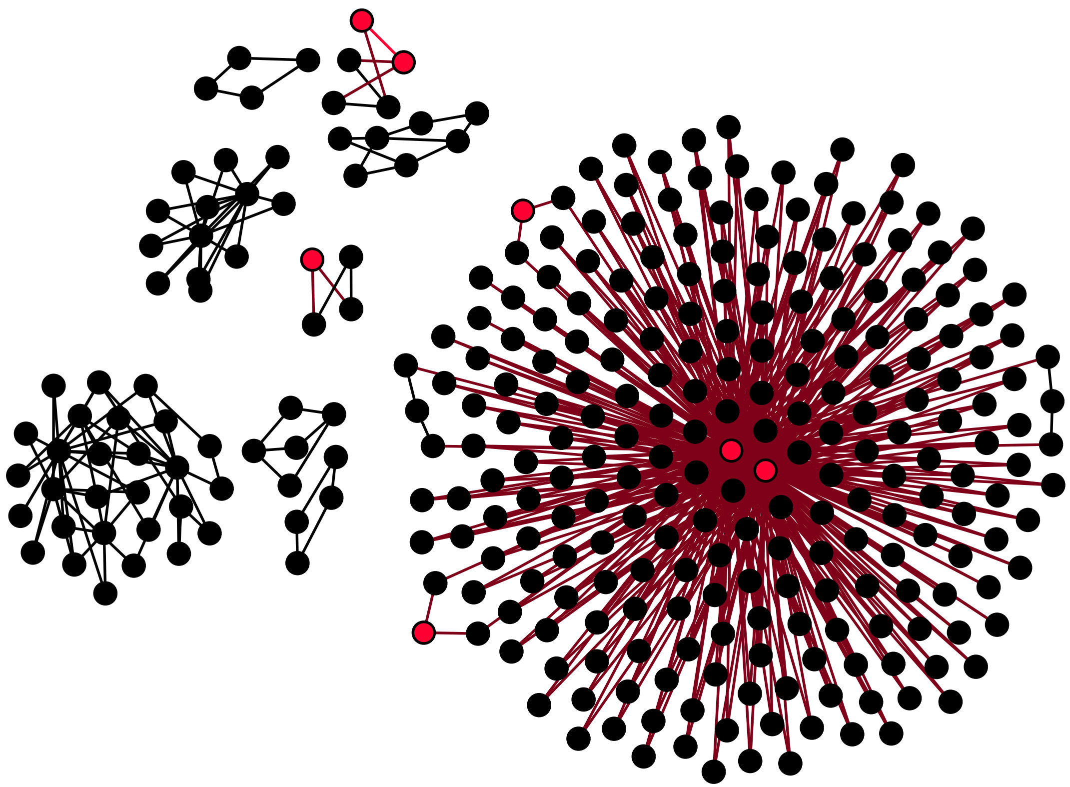
Cliques involving 4 nodes or more



Cliques involving 4 nodes or more



Cliques involving 4 nodes or more



Thank you Peyresq!

Some papers:
Stream graphs: `https://arxiv.org/abs/1710.04073`

Extensions: `https://arxiv.org/abs/1906.04840`

github.com/TiphaineV tiphaine.viard@telecom-paris.fr