Analysis of **Financial Transactions** with **Link Streams**

Matthieu Latapy, Tiphaine Viard, Clémence Magnien

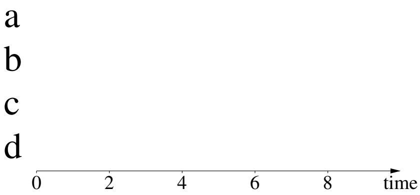
http://complexnetworks.fr

latapy@complexnetworks.fr

LIP6 – CNRS and Sorbonne Université

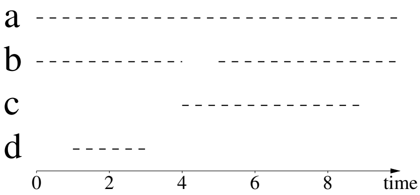
Paris, France

interactions over time



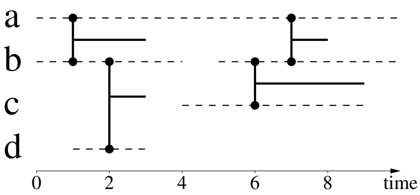
- *a*, *b*, *c*, and *d* for 10 time units

interactions over time



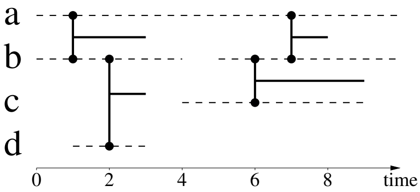
- *a*, *b*, *c*, and *d* for 10 time units
- *a* always present, *b* leaves from 4 to 5, *c* present from 4 to 9, *d* from 1 to 3

interactions over time



- *a*, *b*, *c*, and *d* for 10 time units
- *a* always present, *b* leaves from 4 to 5, *c* present from 4 to 9, *d* from 1 to 3
- *a* and *b* interact from 1 to 3 and from 7 to 8; *b* and *c* from 6 to 9; *b* and *d* from 2 to 3.

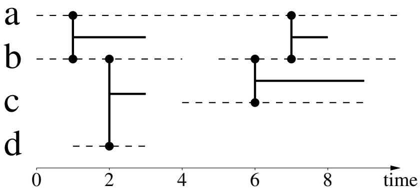
interactions over time



- *a*, *b*, *c*, and *d* for 10 time units
- *a* always present, *b* leaves from 4 to 5, *c* present from 4 to 9, *d* from 1 to 3
- *a* and *b* interact from 1 to 3 and from 7 to 8; *b* and *c* from 6 to 9; *b* and *d* from 2 to 3.

e.g., social interactions, network traffic, money transfers, chemical reactions, etc.

interactions over time

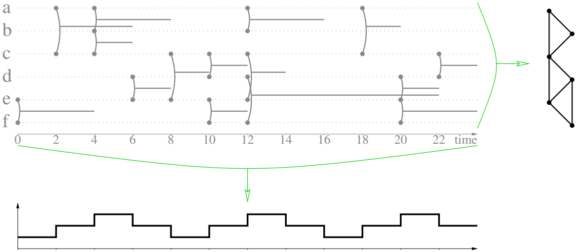


- *a*, *b*, *c*, and *d* for 10 time units
- *a* always present, *b* leaves from 4 to 5, *c* present from 4 to 9, *d* from 1 to 3
- *a* and *b* interact from 1 to 3 and from 7 to 8; *b* and *c* from 6 to 9; *b* and *d* from 2 to 3.

e.g., social interactions, network traffic, money transfers, chemical reactions, etc.

how to describe such data?

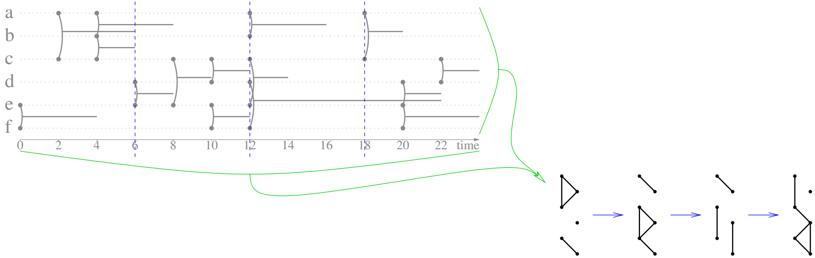
structure or dynamics



graph theory
network science
$\longrightarrow$ structure

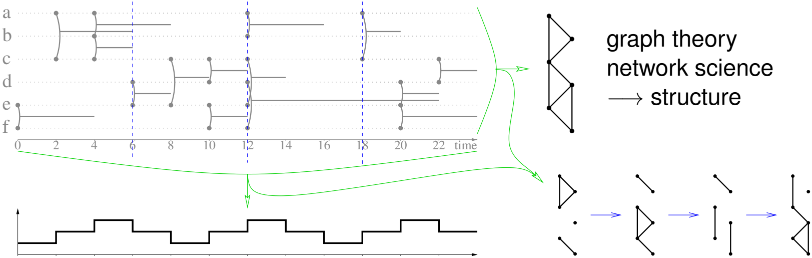
signal analysis, time series $\longrightarrow$ dynamics

structure and dynamics?



time slices
→ graph sequence

structure and dynamics?



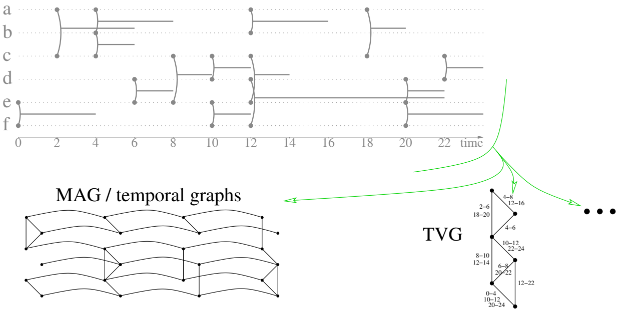
graph theory
network science
⟶ structure

signal analysis, time series ⟶ dynamics

time slices
→ graph sequence

information loss
what slices?
graph sequences?

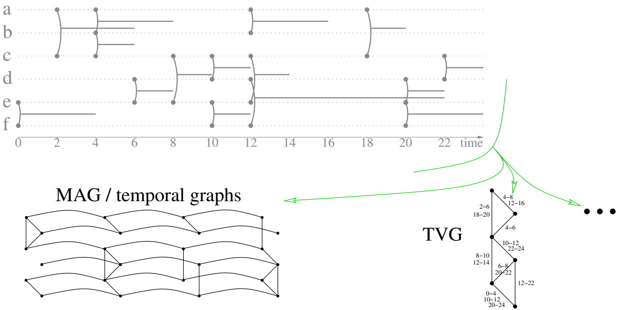
structure and dynamics



lossless but graph-oriented

+ ad-hoc properties (mostly path-related) + contact sequences + relational event models + ...

structure and dynamics



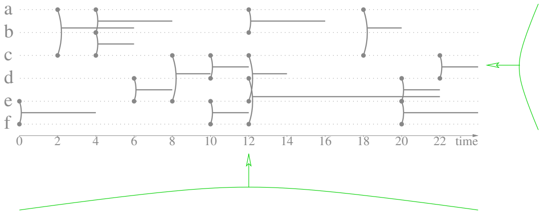
lossless but graph-oriented

+ ad-hoc properties (mostly path-related) + contact sequences + relational event models + ...

what we propose

deal with the stream directly

stream graphs and link streams



graph theory
network science

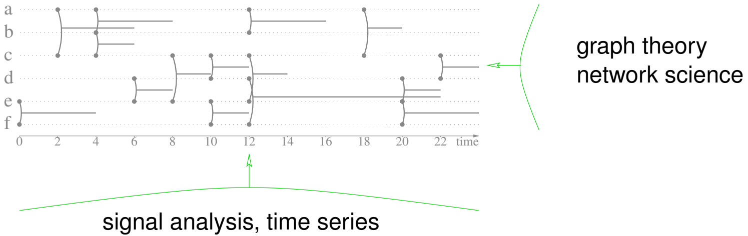
signal analysis, time series

wanted features: simple and intuitive, comprehensive, time-node consistent, generalizes graphs/signal

what we propose

deal with <u>the stream</u> directly

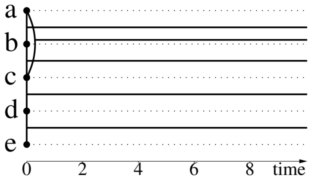
stream graphs and link streams



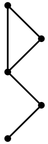
wanted features: simple and intuitive, comprehensive, time-node consistent, generalizes graphs/signal

graph-equivalent streams

stream with no dynamics: nodes always present, either always or never linked $\Longleftrightarrow$ **graph**

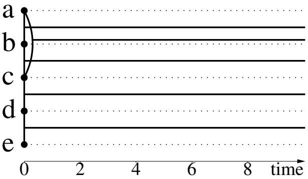


$\Longleftrightarrow$

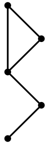


graph-equivalent streams

stream with no dynamics: nodes always present, either always or never linked ⟺ **graph**



⟺



stream properties = **graph properties**

↪ **generalizes graph theory**

our approach

very careful generalization of the most basic concepts

stream graphs and link streams
numbers of nodes and links
clusters and induced sub-streams
density and paths

$\hookrightarrow$ **buliding blocks for higher-level concepts**

neighborhood and degrees
clustering coefficient
betweenness centrality
many others

+ ensure consistency with graph theory

+ ensure classical relations are preserved

definition of stream graphs

Graph $G = (V, E)$ with $E \subseteq V \otimes V$
$uv \in E \;\Leftrightarrow\; u$ and v are linked

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

T: time interval, V: node set
$W \subseteq T \times V, \; E \subseteq T \times V \otimes V$

$(t, v) \in W \;\Leftrightarrow\;$ **v is present at time t**

$$T_v = \{t, (t, v) \in W\}$$

$(t, uv) \in E \;\Leftrightarrow\;$ **u and v are linked at time t**

$$T_{uv} = \{t, (t, uv) \in E\}$$

$(t, uv) \in E$ requires $(t, u) \in W$ and $(t, v) \in W$
i.e. $T_{uv} \subseteq T_u \cap T_v$

definition of stream graphs

Graph $G = (V, E)$ with $E \subseteq V \otimes V$
$uv \in E \;\Leftrightarrow\; u$ and v are linked

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

T: time interval, V: node set
$W \subseteq T \times V, \;\; E \subseteq T \times V \otimes V$

$(t, v) \in W \;\Leftrightarrow\;$ **v is present at time t**

$$T_v = \{t, (t, v) \in W\}$$

$(t, uv) \in E \;\Leftrightarrow\;$ **u and v are linked at time t**

$$T_{uv} = \{t, (t, uv) \in E\}$$

$(t, uv) \in E$ requires $(t, u) \in W$ and $(t, v) \in W$
i.e. $T_{uv} \subseteq T_u \cap T_v$

definition of stream graphs

Graph $G = (V, E)$ with $E \subseteq V \otimes V$
$uv \in E \;\Leftrightarrow\; u$ and v are linked

Stream graph $S = (T, V, W, E)$
T: time interval, V: node set
$W \subseteq T \times V, \;\; E \subseteq T \times V \otimes V$

$(t, v) \in W \;\Leftrightarrow\;$ v **is present at time** t

$$T_v = \{t, (t, v) \in W\}$$

$(t, uv) \in E \;\Leftrightarrow\;$ u **and** v **are linked at time** t

$$T_{uv} = \{t, (t, uv) \in E\}$$

$(t, uv) \in E$ requires $(t, u) \in W$ and $(t, v) \in W$
i.e. $T_{uv} \subseteq T_u \cap T_v$

definition of stream graphs

Graph $G = (V, E)$ with $E \subseteq V \otimes V$
$uv \in E \;\Leftrightarrow\; u$ and v are linked

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

T: time interval, V: node set

$W \subseteq T \times V, \;\; E \subseteq T \times V \otimes V$

$(t, v) \in W \;\Leftrightarrow\;$ **v is present at time t**

$$T_v = \{t, (t, v) \in W\}$$

$(t, uv) \in E \;\Leftrightarrow\;$ **u and v are linked at time t**

$$T_{uv} = \{t, (t, uv) \in E\}$$

$(t, uv) \in E$ requires $(t, u) \in W$ and $(t, v) \in W$

i.e. $T_{uv} \subseteq T_u \cap T_v$

definition of stream graphs

Graph $G = (V, E)$ with $E \subseteq V \otimes V$
$uv \in E \;\Leftrightarrow\; u$ and v are linked

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

T: time interval, V: node set

$W \subseteq T \times V, \;\; E \subseteq T \times V \otimes V$

$(t, v) \in W \;\Leftrightarrow\;$ v **is present at time** t

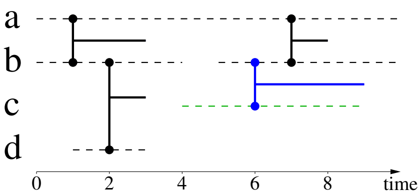
$$T_v = \{t, (t, v) \in W\}$$

$(t, uv) \in E \;\Leftrightarrow\;$ u **and** v **are linked at time** t

$$T_{uv} = \{t, (t, uv) \in E\}$$

$(t, uv) \in E$ requires $(t, u) \in W$ and $(t, v) \in W$
i.e. $T_{uv} \subseteq T_u \cap T_v$

an example



$T = [0, 10]$ $V = \{a, b, c, d\}$

$W = T \times \{a\} \cup ([0, 4] \cup [5, 10]) \times \{b\} \cup \mathbf{[4, 9] \times \{c\}} \cup [1, 3] \times \{d\}$

$T_a = T$ $T_b = [0, 4] \cup [5, 10]$ $\mathbf{T_c = [4, 9]}$ $T_d = [1, 3]$

$E = ([1, 3] \cup [7, 8]) \times \{ab\} \cup \mathbf{[6, 9] \times \{bc\}} \cup [2, 3] \times \{bd\}$

$T_{ab} = [1, 3] \cup [7, 8]$ $\mathbf{T_{bc} = [6, 9]}$ $T_{bd} = [2, 3]$ $T_{ad} = \emptyset$

a few remarks

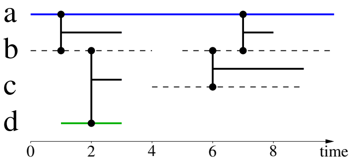
works with... discrete time, continuous time, instantaneous interactions or with durations, directed, weighted, bipartite...

if $\forall v, T_v = T$ then $S \sim L = (T, V, E)$ is a **link stream**

if $\forall u, v, T_{uv} \in \{T, \emptyset\}$ then $S \sim G = (V, E)$ is a **graph-equivalent stream**

size of a stream graph

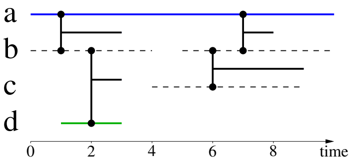
How many nodes? How many links?



$|T_a| = 10 \neq |T_d| = 2$

size of a stream graph

How many nodes? How many links?



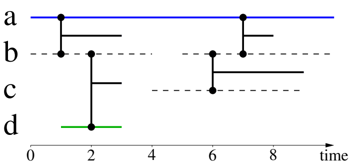
$$|\mathbf{T_a}| = \mathbf{10} \quad \neq \quad |\mathbf{T_d}| = \mathbf{2}$$

$$n = \sum_{v \in V} \frac{|T_v|}{|T|}$$

$$n = \frac{|\mathbf{T_a}|}{\mathbf{10}} + \frac{|T_b|}{10} + \frac{|T_c|}{10} + \frac{|\mathbf{T_d}|}{\mathbf{10}} = \mathbf{1} + 0.9 + 0.5 + \mathbf{0.2} = 2.6 \text{ nodes}$$

size of a stream graph

How many nodes? How many links?



$$|\mathbf{T_a}| = \mathbf{10} \neq |\mathbf{T_d}| = \mathbf{2}$$

$$n = \sum_{v \in V} \frac{|T_v|}{|T|}$$

$$m = \sum_{uv \in V \otimes V} \frac{|T_{uv}|}{|T|}$$

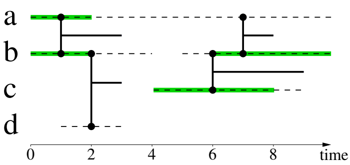
$$n = \frac{|\mathbf{T_a}|}{\mathbf{10}} + \frac{|T_b|}{10} + \frac{|T_c|}{10} + \frac{|\mathbf{T_d}|}{\mathbf{10}} = \mathbf{1} + 0.9 + 0.5 + \mathbf{0.2} = 2.6 \text{ nodes}$$

$$m = \frac{|T_{ab}|}{10} + \frac{|T_{bc}|}{10} + \frac{|T_{bd}|}{10} = 0.3 + 0.3 + 0.1 = 0.7 \text{ links}$$

clusters, sub-streams

Cluster in $G = (V, E)$: a subset of V.

Cluster in $S = (T, V, W, E)$: **a subset of $W \subseteq T \times V$**.



$C = [0,2] \times \{a\} \cup ([0,2] \cup [6,10]) \times \{b\} \cup [4,8] \times \{c\}$

$S(C)$ sub-stream induced by C

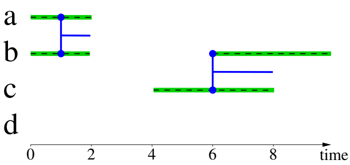
$S(C) = (T, V, \mathbf{C}, \mathbf{E_C})$

$\hookrightarrow$ properties of (sub-streams induced by) clusters

clusters, sub-streams

Cluster in $G = (V, E)$: a subset of V.

Cluster in $S = (T, V, W, E)$: **a subset of $W \subseteq T \times V$**.



$C = [0, 2] \times \{a\} \cup ([0, 2] \cup [6, 10]) \times \{b\} \cup [4, 8] \times \{c\}$

$S(C)$ sub-stream induced by C

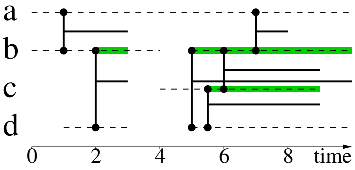
$S(C) = (T, V, \mathbf{C}, \mathbf{E_C})$

$\hookrightarrow$ properties of (sub-streams induced by) clusters

neighborhood

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

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



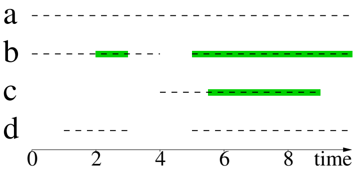
$N(d) = ([2, 3] \cup [5, 10]) \times \{b\} \cup [5.5, 9] \times \{c\}$

$N(v)$ is a cluster

degree

in G and in S:

$d(v)$ is the size of $N(v)$



$$N(d) = ([2,3] \cup [5,10]) \times \{b\} \cup [5.5,9] \times \{c\}$$

$$d(d) = \frac{|[2,3] \cup [5,10]|}{10} + \frac{|[5.5,9]|}{10} = 0.6 + 0.35 = 0.95$$

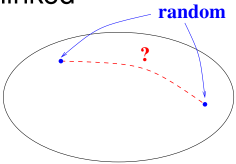
- degree distribution, average degree, etc
- if graph-equivalent stream then graph degree
- relation with n and m

density

in G:

proba two random nodes are linked

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



in S:

proba two random nodes are linked

at a random time instant

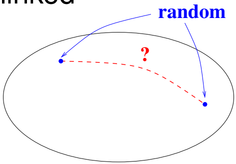
$$\delta(S) = \frac{\textbf{nb links}}{\textbf{nb possible links}} = \frac{\sum_{uv \in V \otimes V} |T_{uv}|}{\sum_{uv \in V \otimes V} |T_u \cap T_v|}$$

density

in G:

proba two random nodes are linked

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

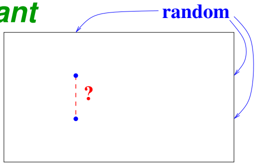


in S:

proba two random nodes are linked

at a random time instant

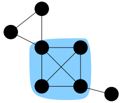
$$\delta(S) = \frac{\textbf{nb links}}{\textbf{nb possible links}} = \frac{\sum_{uv \in V \otimes V} |T_{uv}|}{\sum_{uv \in V \otimes V} |T_u \cap T_v|}$$



- if graph-equivalent stream then graph density
- relation with n, m, and average degree

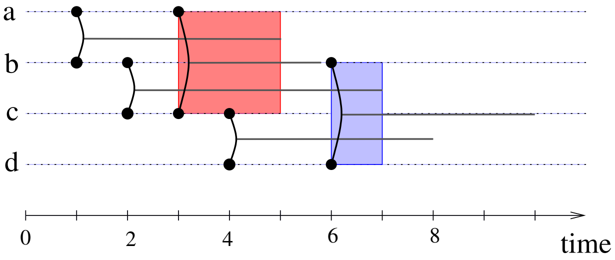
cliques

in G: sub-graph of density 1
all nodes are linked together



in S: **sub-stream of density 1**

all nodes interact all the time

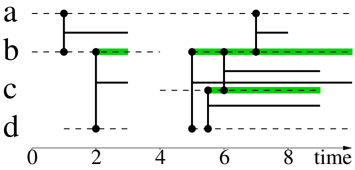


clustering coefficient

in G and in S:

density of the neighborhood

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



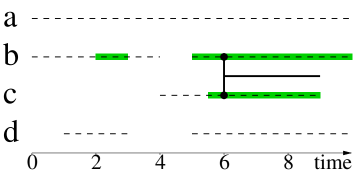
$$N(d) = ([2,3] \cup [5,10]) \times \{b\} \cup [5.5,9] \times \{c\}$$

clustering coefficient

in G and in S:

density of the neighborhood

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

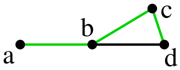


$$N(d) = ([2,3] \cup [5,10]) \times \{b\} \cup [5.5,9] \times \{c\}$$

$$cc(d) = \delta(N(d)) = \frac{|[6,9]|}{|[5.5,9]|} = \frac{6}{7}$$

paths

in G:



from a to d:
(a, b), (b, c), (c, d)
length: 3

→ shortest paths

in S:

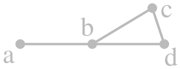
from $(1, d)$ to $(9, c)$:
$(2, d, b)$, $(3, b, a)$, $(7.5, a, b)$, $(8, b, c)$

length: 4
duration: 6

→ shortest paths
→ fastest paths

paths

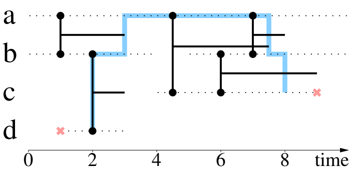
in *G*:



from *a* to *d*:
$(a, b), (b, c), (c, d)$
length: 3

→ shortest paths

in *S*:



from $(1, d)$ to $(9, c)$:

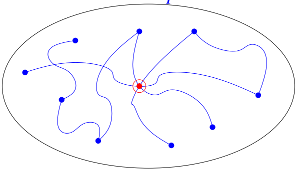
$(2, d, b), (3, b, a), (7.5, a, b), (8, b, c)$

length: 4
duration: 6

→ shortest paths
→ fastest paths

betweenness centrality

in G:
$b(v)$ = fraction of
shortest paths
from any u to any w in V
that involve v



in S:
$b(t, v)$ = fraction of
~~*shortest*~~ **fastest** *paths*
from any (i, u) to any (j, w) in W
that involve (t, v)

betweenness centrality

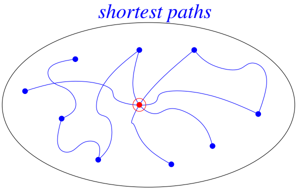
in G:

$b(v)$ = fraction of

shortest paths

from any u to any w in V

that involve v



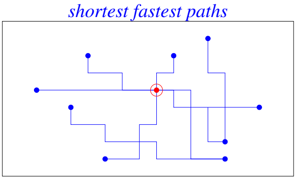
in S:

$b(t, v)$ = fraction of

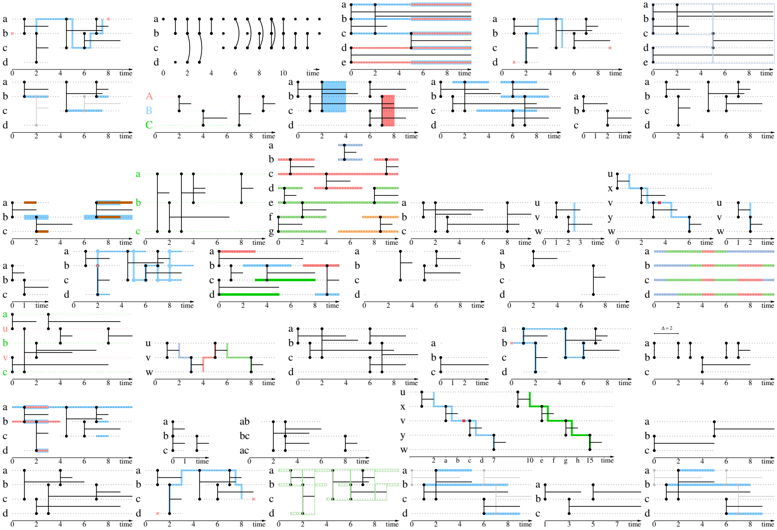
shortest **fastest** *paths*

from any (i, u) to any (j, w) in W

that involve (t, v)



many other concepts



arxiv preprint – SNAM publication

conclusion

we provide a language (set of concepts) that:

- makes it easy to deal with interaction traces,
- combines structure and dynamics in a consistent way,
- generalizes graphs / networks ; **signals / time series ?**
- meets classical and **new algorithmic challenges**,
- opens new **perspectives for data analysis**,
- clarifies the interplay **interactions** $\longleftrightarrow$ **relations**.

studies in progress: internet traffic, financial transactions, mobility/contacts, mailing-lists, sales, etc.