Stream Graphs, Link Streams and Related Algorithmic Challenges

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.

\[ e.g., \text{social interactions, network traffic, money transfers, chemical reactions, etc.} \]

how to describe such data?
structure or dynamics

graph theory
network science
→ structure

signal analysis, time series → dynamics
structure and dynamics?

Graph theory
→ structure
network science
→ dynamics

time slices
→ graph sequence

Context
Approach
Basics
Degrees
Density
Paths
Further
Algorithms
structure and dynamics?

signal analysis, time series $\rightarrow$ dynamics

graph theory
network science
$\rightarrow$ structure

time slices
$\rightarrow$ 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

MAG / temporal graphs

TVG

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

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

deal with the stream 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

\[ a \rightarrow b \rightarrow c \rightarrow d \rightarrow e \]

\[ 0 \rightarrow 2 \rightarrow 4 \rightarrow 6 \rightarrow 8 \text{ time} \]

\[ \iff \]

graph

\[ \rightarrow \]
graph-equivalent streams

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

\[
\begin{array}{cccccc}
0 & 2 & 4 & 6 & 8 & \text{time} \\
\hline
a & \cdots & \cdots & \cdots & \cdots & \\
b & \cdots & \cdots & \cdots & \cdots & \\
c & \cdots & \cdots & \cdots & \cdots & \\
d & \cdots & \cdots & \cdots & \cdots & \\
e & \cdots & \cdots & \cdots & \cdots & \\
\end{array}
\]

\[\iff\]

graph

\[\iff\]

stream properties \quad = \quad \text{graph properties}

\[\iff\]

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

building 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 $E \subseteq V \otimes V$

$uv \in E \iff u \text{ and } v \text{ are linked}$

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

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

$(t, v) \in W \iff v \text{ is present at time } t$

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

$(t, uv) \in E \iff u \text{ and } v \text{ are linked at time } t$

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

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

i.e. $T_{uv} \subseteq T_u \cap T_v$
**definition of stream graphs**

Graph $G = (V, E)$ with $E \subseteq V \otimes V$

$uv \in E \iff u$ and $v$ are linked

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

$T$: time interval, $V$: node set

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

$(t, v) \in W \iff v$ is present at time $t$

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

$(t, uv) \in E \iff u$ 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 $E \subseteq V \otimes V$

$uv \in E \iff u$ and $v$ are linked

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

$T$: time interval, $V$: node set

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

$(t, v) \in W \iff v$ is present at time $t$

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

$(t, uv) \in E \iff u$ 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 $E \subseteq V \otimes V$

$uv \in E \iff u$ and $v$ are linked

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

$T$: time interval, $V$: node set

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

$(t, v) \in W \iff v$ is present at time $t$

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

$(t, uv) \in E \iff u$ 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 $E \subseteq V \otimes V$

$uv \in E \iff u$ and $v$ are linked

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

$T$: time interval, $V$: node set

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

$(t, v) \in W \iff v$ is present at time $t$

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

$(t, uv) \in E \iff u$ 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

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

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

\[ T_a = T \quad T_b = [0, 4] \cup [5, 10] \quad T_c = [4, 9] \quad T_d = [1, 3] \]

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

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

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

\[ n = \frac{|T_a|}{10} + \frac{|T_b|}{10} + \frac{|T_c|}{10} + \frac{|T_d|}{10} = 1 + 0.9 + 0.5 + 0.2 = 2.6 \text{ nodes} \]
size of a stream graph

How many nodes? How many links?

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

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

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

\[ m = \frac{|T_{ab}|}{10} + \frac{|T_{bc}|}{10} + \frac{|T_{bd}|}{10} = 0.3 + 0.3 + 0.1 = 0.7 \text{ links} \]
clusters, sub-streams

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

\begin{align*}
C &= [0, 2] \times \{a\} \cup ([0, 2] \cup [6, 10]) \times \{b\} \cup [4, 8] \times \{c\} \\
S(C) &= \text{sub-stream induced by } C \\
S(C) &= (T, V, C, E_C)
\end{align*}

$\rightarrow$ 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 $C$
$S(C) = (T, V, C, E_C)$

→ properties of (sub-streams induced by) clusters
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
in $G$ and in $S$:

\[ d(v) \text{ is the size of } N(v) \]

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

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

- 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

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

• if graph-equivalent stream then graph density
• relation with \( n, m, \) and average degree
Cliquets

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

\[ cc(v) = \delta(N(v)) \]

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

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

$$b(v) = \text{fraction of shortest paths from any } u \text{ to any } w \text{ in } V \text{ that involve } v$$

in $S$:

$$b(t, v) = \text{fraction of shortest fastest paths from any } (i, u) \text{ to any } (j, w) \text{ in } W \text{ that involve } (t, v)$$
betweenness centrality

in $G$:

$$b(v) = \text{fraction of shortest paths from any } u \text{ to any } w \text{ in } V \text{ that involve } v$$

in $S$:

$$b(t, v) = \text{fraction of shortest fastest paths from any } (i, u) \text{ to any } (j, w) \text{ in } W \text{ that involve } (t, v)$$
many other concepts
algorithmic concerns

extension of graph concepts...

...extension of graph algorithms?

some properties of $S$ derive from properties of $G_t$

- neighborhood
- degrees
- $k$-cores

some don’t but algorithms may be adapted

- density
- cliques (greedy, Bron-Kerbosch)

some still don’t $\Rightarrow$ new algorithms needed

- (directed) paths
- betweenness
- patterns
algorithmic concerns

extension of graph concepts...
...extension of graph algorithms?

some properties of $S$ derive from properties of $G_t$
- neighborhood, degrees, $k$-cores, ...

some don’t but algorithms may be adapted
- density, cliques (greedy, Bron-Kerbosch), ...

some still don’t $\Rightarrow$ new algorithms needed
- (directed) paths, betweenness, patterns, ...
**algorithmic challenges**

- **classical ones**
  - streaming/on-line
  - fully dynamic
  - approximation
  - space complexity

- **new ones**
  - cliques, paths, betweenness
  - unbounded number of links
  - prediction?

- **good news**
  - time-induced locality
  - knowledge of dynamics
  - better than induced graph?
algorithmic challenges

classical ones
streaming/on-line
fully dynamic
approximation
space complexity

new ones
cliques, paths, betweenness
unbounded number of links
prediction?

good news
time-induced locality
knowledge of dynamics
better than induced graph?
algorithmic challenges

classical ones

streaming/on-line
fully dynamic
approximation
space complexity

new ones

cliques, paths, betweenness
unbounded number of links
prediction?

good news

time-induced locality
knowledge of dynamics
better than induced graph?
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 $\leftrightarrow$ relations.

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

special issues of international journals

Theoretical Computer Science (TCS)

Link Streams: models and algorithms

Computer Networks

Link Streams: methods and case studies

deadline: July 1st

http://link-streams.com