



HAL
open science

Flots de liens et stream graphs pour la modélisation des interactions temporelles

Matthieu Latapy, Tiphaine Viard, Clémence Magnien

► To cite this version:

Matthieu Latapy, Tiphaine Viard, Clémence Magnien. Flots de liens et stream graphs pour la modélisation des interactions temporelles. ALGOTEL 2018 - 20èmes Rencontres Francophones sur les Aspects Algorithmiques des Télécommunications, May 2018, Roscoff, France. hal-01782899

HAL Id: hal-01782899

<https://hal.science/hal-01782899v1>

Submitted on 2 May 2018

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Flots de liens et stream graphs pour la modélisation des interactions temporelles[†]

Matthieu Latapy, Tiphaine Viard et Clémence Magnien

Sorbonne Université, CNRS, Laboratoire d'Informatique de Paris 6, LIP6, F-75005 Paris, France

La plupart des livres de référence sur les graphes commencent par introduire le même ensemble de concepts élémentaires, puis définissent des concepts plus avancés sur cette base. Nous montrons ici que cette approche peut être généralisée pour traiter non seulement la structure mais aussi la dynamique, de façon unifiée. On obtient un langage pour décrire les interactions temporelles, similaire au langage fourni par la théorie des graphes pour décrire les relations.

Mots-clés : stream graphs, link streams, temporal networks, time-varying graphs, dynamic graphs, dynamic networks, longitudinal networks, interactions, time, graphs, networks

Basic definitions

Given two sets X and Y , we denote by $X \times Y$ the set of ordered pairs (x, y) with $x \in X$ and $y \in Y$, and we denote by $X \otimes Y$ the set of unordered pairs of distinct elements $xy = yx$ with $x \in X$, $y \in Y$, and $x \neq y$. Then, a graph $G = (V, E)$ is classically defined by a set of nodes V and a set of links $E \subseteq V \otimes V$.

We define a **stream graph** $S = (T, V, W, E)$ by a finite set of nodes V , a continuous time interval $T \subset \mathbb{R}$, a set of temporal nodes $W \subseteq T \times V$, and a set of links $E \subseteq T \times V \otimes V$ such that $(t, uv) \in E$ implies $(t, u) \in W$ and $(t, v) \in W$.

We denote by T_v the set of time instants at which v is present, by T_{uv} the set of time instants at which uv is present, by V_t the set of nodes present at time t , and by E_t the set of links present at time t : $T_v = \{t, (t, v) \in W\}$, $T_{uv} = \{t, (t, uv) \in E\}$, $V_t = \{v, (t, v) \in W\}$, and $E_t = \{uv, (t, uv) \in E\}$. Notice that $T_{uv} \subseteq T_u \cap T_v$.

Stream graphs model interactions between nodes over time, as well as the dynamics of nodes themselves. For instance, nodes may represent individuals present in a given building and links may represent contacts between them. Nodes may represent on-line computers and links may represent data exchanges between them. Such examples are countless.

Induced graph, link stream, and graph-equivalent streams

We define $G(S) = (\{v, T_v \neq \emptyset\}, \{uv, T_{uv} \neq \emptyset\}) = (\bigcup_{t \in T} V_t, \bigcup_{t \in T} E_t)$ the **graph induced by S** . It is the graph where there is a link between two nodes if they interacted at least once.

If all nodes are present all the time, *i.e.* $T_v = T$ for all v or, equivalently, $V_t = V$ for all t , then we say that S is a **link stream** and we denote it by $L = (T, V, E)$ (with $W = T \times V$ implicitly). Indeed, there is no dynamics on nodes in this case, and S is fully defined by this triplet.

If in addition for all u and v in V , $T_{uv} \in \{\emptyset, T\}$, *i.e.* all existing links are present all the time, then there is no significant distinction between S and $G(S)$, and we say that S is a **graph-equivalent stream**. This gives a formal ground to the fact that stream graphs generalize graphs : we extend graph concepts to stream graphs in a way such that, if a stream graph has a given stream graph property and happens to be a graph-equivalent stream, then this graph has the corresponding graph property.

Example and drawings

We illustrate our definitions with **drawings** like in Figure 1. We display node names on a vertical axis on the left, and time on an horizontal axis at the bottom of the drawing. Each node presence times are

[†]Cet article est un extrait condensé de l'article [3], en cours d'évaluation. Le résumé fourni ici joue aussi le rôle d'introduction.

represented by an horizontal dotted line in front of its name, whenever the node is present. Each link presence times are represented by an horizontal solid line parallel to the two dotted lines of involved nodes, and a vertical solid line joining these two dotted lines (marked with bullets) when the two nodes start interacting.

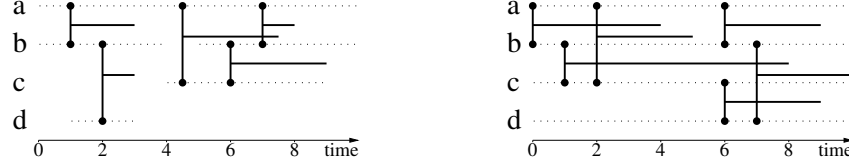


FIGURE 1: Simple examples of stream graphs and link streams. Left : a stream graph $S = (T, V, W, E)$ with $T = [0, 10] \subseteq \mathbb{R}$, $V = \{a, b, c, d\}$, $W = [0, 10] \times \{a\} \cup ([0, 4] \cup [5, 10]) \times \{b\} \cup [4, 9] \times \{c\} \cup [1, 3] \times \{d\}$, and $E = ([1, 3] \cup [7, 8]) \times \{ab\} \cup [4.5, 7.5] \times \{ac\} \cup [6, 9] \times \{bc\} \cup [2, 3] \times \{bd\}$. In other words, $T_a = [0, 10]$, $T_b = [0, 4] \cup [5, 10]$, $T_c = [4, 9]$, $T_d = [1, 3]$, $T_{ab} = [1, 3] \cup [7, 8]$, $T_{ac} = [4.5, 7.5]$, $T_{bc} = [6, 9]$, $T_{bd} = [2, 3]$, and $T_{ad} = T_{cd} = \emptyset$. **Right :** a link stream $L = (T, V, E)$ with $T = [0, 10] \subseteq \mathbb{R}$, $V = \{a, b, c, d\}$, and $E = ([0, 4] \cup [6, 9]) \times \{ab\} \cup [2, 5] \times \{ac\} \cup [1, 8] \times \{bc\} \cup [7, 10] \times \{bd\} \cup [6, 9] \times \{cd\}$. In other words, $T_a = T_b = T_c = T_d = T$ and $T_{ab} = [0, 4] \cup [6, 9]$, $T_{ac} = [2, 5]$, $T_{bc} = [1, 8]$, $T_{bd} = [7, 10]$ and $T_{cd} = [6, 9]$.

Size and duration

In order to capture the fact that some nodes may be present for much longer than others, we define the **contribution of node** v as $n_v = \frac{|T_v|}{|T|}$. We define similarly the contribution of a pair of nodes uv as $m_{uv} = \frac{|T_{uv}|}{|T|}$. We then define the **number of nodes** n and the **number of links** m in S :

$$n = \sum_{v \in V} n_v = \frac{|W|}{|T|} \quad \text{and} \quad m = \sum_{uv \in V \otimes V} m_{uv} = \frac{|E|}{|T|}.$$

Then, each node contributes to the total number of nodes proportionally to its involvement in S : v in V accounts for 1 node only if it is present in S all the time. The same holds for links.

In Figure 1, we obtain for S the values $n = \frac{|T_a|}{10} + \frac{|T_b|}{10} + \frac{|T_c|}{10} + \frac{|T_d|}{10} = 2.6$ nodes and $m = 1$ link. For L , we obtain $n = 4$ nodes and $m = 2.3$ links.

In a link stream $L = (T, V, E)$, $n_v = 1$ for all v and so $n = |V|$. In a graph-equivalent stream, the number of nodes and links in the stream are equal to the number of nodes and links in the corresponding graph.

Density

Since the density of graph $G = (V, E)$ is the probability when one takes a random element uv in $V \otimes V$ that there is a link between u and v in E , we define the **density** of stream graph $S = (T, V, W, E)$ as the probability when one takes a random element (t, uv) of $T \times V \otimes V$ such that (t, u) and (t, v) are in W , that (t, uv) is in E :

$$\delta(S) = \frac{\sum_{uv \in V \otimes V} |T_{uv}|}{\sum_{uv \in V \otimes V} |T_u \cap T_v|} = \frac{\int_{t \in T} |E_t| dt}{\int_{t \in T} |V_t \otimes V_t| dt}$$

The density is the fraction of possible links that do exist, both for graphs and stream graphs.

Notice that $\sum_{uv \in V \otimes V} |T_{uv}| = \int_{t \in T} |E_t| dt = |E|$.

For S defined in Figure 1 (left), we obtain $\delta(S) = \frac{10}{22} \sim 0.45$, and for L defined in this figure (right), $\delta(L) = \frac{23}{60} \sim 0.38$.

There is in general no relation between the density δ , the number of nodes n and the number of links m in a stream graph. However, the classical graph relation $\delta = \frac{2m}{n(n-1)}$ holds for link streams. In addition, the density of a graph-equivalent stream is equal to the density of the corresponding graph.

Substreams and clusters

A graph $G' = (V', E')$ is a subgraph of $G = (V, E)$ if $V' \subseteq V$ and $E' \subseteq E$. A cluster C of $G = (V, E)$ is a subset of V . The set of links between nodes in C is $E(C) = \{uv \in E, u \in C \text{ and } v \in C\}$, and $G(C) = (C, E(C))$ denotes the subgraph of G induced by C .

We say that a stream $S' = (T', V', W', E')$ is a **substream** of $S = (T, V, W, E)$ if $T' \subseteq T$, $V' \subseteq V$, $W' \subseteq W$, and $E' \subseteq E$. We define a **cluster** C of $S = (T, V, W, E)$ as a subset of W . We define the set of links between nodes involved in C as $E(C) = \{(t, uv) \in E, (t, u) \in C \text{ and } (t, v) \in C\}$, and we denote by $S(C) = (T, V, C, E(C))$ the **substream of S induced by C** . See Figure 2.

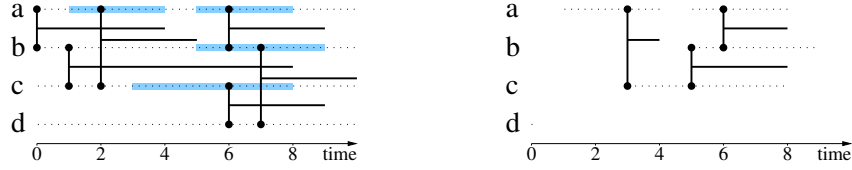


FIGURE 2: An example of cluster with its induced substream. Left : the cluster, displayed in blue, is $C = ([1, 4] \cup [5, 8]) \times \{a\} \cup [5, 9] \times \{b\} \cup [3, 8] \times \{c\}$. Right : the substream induced by C is $S(C) = ([0, 10], \{a, b, c, d\}, C, E(C))$ with $E(C) = [6, 8] \times \{ab\} \cup [3, 4] \times \{ac\} \cup [5, 8] \times \{bc\}$.

Cliques

Just like in graphs, a **clique** of stream graph S is a cluster C of S of density 1. In other words, all pairs of nodes involved in C are linked in S whenever both are involved in C . A clique C is **maximal** if there is no other clique C' such that $C \subset C'$, and a clique C is **compact** if there exists $[b, e] \subseteq T$ and $X \subseteq V$ such that $C = [b, e] \times X$. See Figure 3 for an illustration.

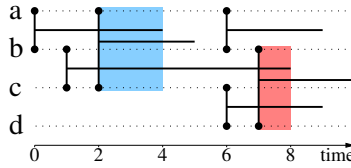


FIGURE 3: Examples of maximal compact cliques. We display the two maximal compact cliques involving three nodes of the link stream L of Figure 1 (right) : $[2, 4] \times \{a, b, c\}$ and $[7, 8] \times \{b, c, d\}$. Its other maximal compact cliques are $[0, 4] \times \{a, b\}$, $[6, 9] \times \{a, b\}$, $[2, 5] \times \{a, c\}$, $[1, 8] \times \{b, c\}$, $[7, 10] \times \{b, d\}$, $[6, 9] \times \{c, d\}$ (involving two nodes each).

If one considers a graph-equivalent stream, then its maximal cliques are necessarily compact, and they correspond exactly to the maximal cliques of its induced graph.

Neighborhood and degree

In the graph $G = (V, E)$, the neighborhood $N(v)$ of $v \in V$ is the cluster $N(v) = \{u, uv \in E\}$, and the degree $d(v)$ of v is the number of nodes in this cluster, which is equal to the number of links involving v . We then have $\sum_{v \in V} d(v) = 2 \cdot m$.

In a stream graph $S = (T, V, W, E)$, we define the **neighborhood of a node v** as the cluster $N(v) = \{(t, u), (t, uv) \in E\}$ and the **degree $d(v)$ of v** as the number of nodes in this cluster. As with graphs, this is equal to the number of links involving v : $d(v) = \frac{|N(v)|}{|T|} = \sum_{u \in V} \frac{|T_{uv}|}{|T|} = \sum_{u \in V} m_{uv}$. See Figure 4 for an illustration.

As with graphs, the sum of the degree of all nodes in S is equal to twice the number of links in S : $\sum_{v \in V} d(v) = \sum_{v \in V} \sum_{u \in V} \frac{|T_{uv}|}{|T|} = 2 \cdot m$. In addition, if we consider a graph-equivalent stream, then the degree of any of its nodes is equal to the degree of this node in the corresponding graph.

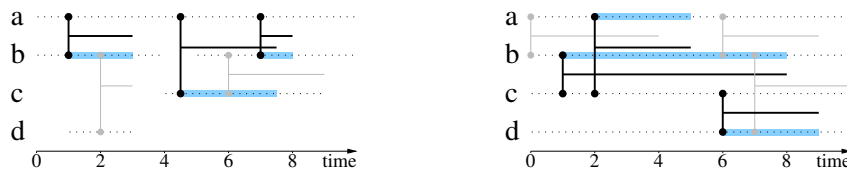


FIGURE 4: Two examples of neighborhoods and degrees of nodes. We display in black the links involving the node under concern, and in grey the other links. Left : $N(a) = ([1, 3] \cup [7, 8]) \times \{b\} \cup [4.5, 7.5] \times \{c\}$ is in blue, leading to $d(a) = \frac{3}{10} + \frac{3}{10} = 0.6$. Right : $N(c) = [2, 5] \times \{a\} \cup [1, 8] \times \{b\} \cup [6, 9] \times \{d\}$ is in blue, leading to $d(c) = \frac{13}{10} = 1.3$.

Clustering coefficient

Just like in graphs, the **clustering coefficient** of a given node v is the density of its neighborhood : $cc(v) = \delta(N(v))$, leading to :

$$cc(v) = \delta(N(v)) = \frac{\sum_{uw \in V \otimes V} |T_{vu} \cap T_{vw} \cap T_{uw}|}{\sum_{uw \in V \otimes V} |T_{vu} \cap T_{vw}|}$$

In other words, in a graph $cc(v)$ is the probability that two randomly chosen neighbors of v are linked together in G ; in a stream graph, $cc(v)$ is the probability when one takes two random neighbors u and w of v at time t , *i.e.* a random (t, uw) in $T \times V \otimes V$ such that (t, vu) and (t, vw) are in E , that u is linked to w in S at time t , *i.e.* that (t, uw) is in E .

In a graph-equivalent stream, $cc(v)$ in the stream is equal to $cc(v)$ in the corresponding graph.

Conclusion

In this paper, we present some key notions of the formalism we introduced in [3] for **dealing directly with the both temporal and structural nature of interactions over time**. Similarly to graph theory, the strength of our approach is to rely on very basic innovations that make it trivial to define more advanced objects. In addition, our formalism is **consistent with graph theory** : if one considers a stream graph with no dynamics (nodes are present all the time, and two nodes are either linked all the time or not at all), then the stream graph is equivalent to a graph and its stream properties are equivalent to the properties of the corresponding graph. As a consequence, our formalism is a generalization of graph theory, which provides a solid ground for generalizing other graph notions.

Complementary to previous work on complexity classes in time-varying graphs [1], our formalism raises **new algorithmic challenges**, like clique computations in stream graphs for instance [4, 2, 5]. It also opens the way to **better understanding of many real-world data**, like network traffic, mobility traces, and financial transactions, on which we are currently working.

*

Références

- [1] Arnaud Casteigts, Paola Flocchini, Walter Quattrociocchi, and Nicola Santoro. Time-varying graphs and dynamic networks. *IJPEDES*, 27(5) :387–408, 2012.
- [2] Anne-Sophie Himmel, Hendrik Molter, Rolf Niedermeier, and Manuel Sorge. Adapting the bronkerbosch algorithm for enumerating maximal cliques in temporal graphs. *Social Netw. Analys. Mining*, 7(1) :35 :1–35 :16, 2017.
- [3] Matthieu Latapy, Tiphaine Viard, and Clémence Magnien. Stream graphs and link streams for the modeling of interactions over time. *CoRR*, abs/1710.04073, 2017.
- [4] Tiphaine Viard, Matthieu Latapy, and Clémence Magnien. Computing maximal cliques in link streams. *Theor. Comput. Sci.*, 609 :245–252, 2016.
- [5] Tiphaine Viard, Clémence Magnien, and Matthieu Latapy. Enumerating maximal cliques in link streams with durations. *CoRR*, abs/1712.06970, 2017. To appear in Information Processing Letters.