

Les Flots de Liens comme généralisation des Graphes et des Séries Temporelles

Matthieu LATAPY, Tiphaine VIARD, Clémence MAGNIEN

LIP6, CNRS et Sorbonne Université
4 place Jussieu, 75005 Paris, France
Matthieu.Latapy@lip6.fr

Résumé – Les flots de liens sont des séquences de triplets (t, u, v) signifiant que u et v ont interagi à l’instant t . Ces objets, qui apparaissent dans de nombreux contextes, sont souvent étudiés comme des graphes (on oublie la dimension temporelle) ou comme des séries temporelles (on oublie la composante structurelle). A l’inverse, nous proposons ici de les prendre comme objets d’étude à part entière, et de développer un formalisme capable de capturer leur nature à la fois temporelle *et* structurelle. Pour ce faire, nous remarquons que les graphes et les séries temporelles peuvent être les uns comme les autres vus comme des flots de liens particuliers (sans dynamique ou sans structure). Par conséquent, nous proposons de généraliser et unifier la théorie des graphes et celle des séries temporelles, en une théorie des flots de liens. Nous présentons notre généralisation de la théorie des graphes et les questions soulevées par la généralisation des séries temporelles, qui reste à mener.

Abstract – Link streams are sequences of triplets (t, u, v) meaning that u and v interacted at time t . Such objects appear in many contexts, and are often studied as graphs (time information is discarded) or time series (structural information is discarded). Instead, we propose here to consider link streams as objects of study in themselves, and to design a formalism able to capture their intrinsically structural *and* temporal nature. To do this, we notice that both graphs and time series may be seen as particular link streams (with no dynamics or no structure). Therefore, we propose to generalize and unify time series and graph theories into a theory of link streams. We present our graph theory generalization and questions raised by time series generalization, that remains to be done.

1 Contexte

On rencontre dans de très nombreux contextes des séquences de liens comme celle représentée Figure 1. Nous appelons de telles séquences des *flots de liens*; chaque lien est un triplet (t, u, v) signifiant qu’une interaction a eu lieu entre les nœuds u et v à l’instant t . Ces flots de liens peuvent par exemple modéliser du trafic réseau, des mobilités individuelles, ou des transactions financières : (t, u, v) peut signifier que l’ordinateur u a envoyé des données à l’ordinateur v à l’instant t ; qu’un individu u a fréquenté un lieu v à l’instant t ; ou que de l’argent a été transféré du compte u au compte v à l’instant t .

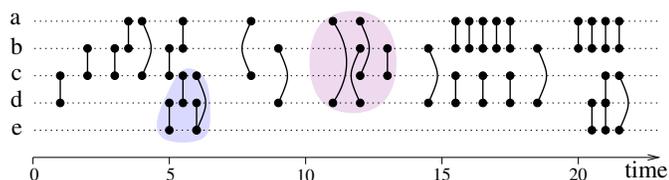


FIGURE 1 – Un exemple de flot de liens représentant des interactions entre a , b , c , d , et e (axe vertical à gauche) entre l’instant 0 et l’instant 22 (axe horizontal en bas). Une interaction a eu lieu entre c et d à l’instant 1, entre b et c aux instants 2 et 3, et ainsi de suite. Des groupes d’interactions comme celles des surfaces colorées peuvent être la signature d’événements intéressants à détecter.

Tirer parti de l’information extrêmement riche capturée par les flots de liens est porteur d’un grand potentiel de progrès pour de nombreux problèmes. Un exemple type est la détection d’événements caractérisés par des séquences spécifiques d’interactions dans un petit groupe de nœuds, pendant un temps limité. Toutefois, comme illustré Figure 2, les approches classiques consistent essentiellement en la transformation des flots de liens en diverses sortes de graphes ou de séries temporelles. Ceci permet de tirer parti des vastes ensembles de méthodes et outils disponible pour ces objets, mais ils capturent mal la nature à la fois temporelle *et* structurelle des flots de liens, et ils induisent d’importantes pertes d’information.

De nombreux efforts sont donc menés pour enrichir ces approches. Par exemple, on découpe le temps en tranches et on construit des séquences de graphes (un pour chaque tranche), mais étudier ces séquences est loin d’être trivial [1, 4]. De plus, choisir des tranches pertinentes est un défi en soi, qui oblige à des compromis peu satisfaisants et d’importantes pertes d’information [4, 6]. Pour aller plus loin, on peut coder toute l’information dans des graphes enrichis de divers types, comme les *Time-Varying Graph* (TVG) : chaque lien du graphe est étiqueté par ses temps de présence dans le flot [2].

Nous proposons une approche orthogonale : plutôt que transformer les flots de liens en graphes ou séries temporelles de diverses sortes, **nous proposons de généraliser et unifier la théorie des graphes et celle des séries temporelles en un for-**

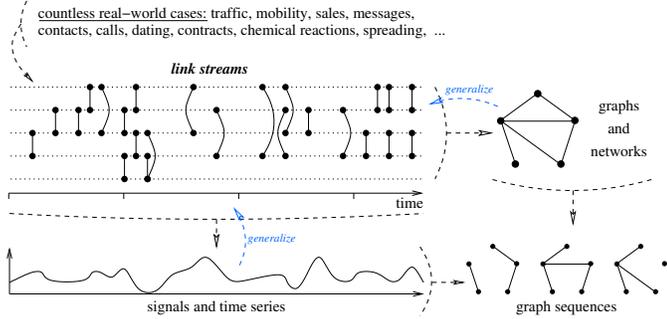


FIGURE 2 – Les approches classiques transforment les flots de liens en graphes, séries temporelles, ou séquences de graphes. Au contraire, nous proposons de généraliser théories des graphes et des séries temporelles en un formalisme dédié aux flots de liens.

malisme dédié aux flots de liens. En effet, comme nous allons le voir, les graphes et les séries temporelles peuvent être vus comme des cas particuliers de flots de liens. Par conséquent, les propriétés usuelles définies sur les graphes et les séries temporelles doivent pouvoir être généralisées en des concepts définis sur les flots de liens. Nous espérons obtenir ainsi un formalisme capable d’appréhender pleinement la nature à la fois temporelle et structurelle des flots de liens, et tirant le meilleur parti des deux mondes.

Nous présentons d’abord le cadre général que nous proposons, puis quelques éléments clés de notre généralisation de la théorie des graphes, récemment publiée [3]. Nous montrerons ensuite comment les séries temporelles peuvent également être vues comme des flots de liens particuliers. La généralisation des concepts usuels sur les séries temporelles reste à mener, et nous pointerons les questions qui semblent les plus importants à ce stade.

2 Le cadre général

Dans toute la suite, étant donné un ensemble X , nous notons $X \otimes X$ l’ensemble des paires d’éléments distincts a et b de X , que nous notons ab . Les paires ne sont pas ordonnées : $ab = ba$.

Comme illustré Figure 3, on définit un **stream graph** $S = (T, V, W, E)$ par un intervalle T de \mathbb{R} représentant le temps, un ensemble fini V de nœuds, un ensemble $W \subseteq T \times V$ de nœuds temporels, et un ensemble $E \subseteq T \times V \otimes V$ de liens tels que $(t, uv) \in E$ implique $(t, u) \in W$ et $(t, v) \in W$.

On définit $v_t = 1$ si $(t, v) \in W$ et $v_t = 0$ sinon, ainsi que $uv_t = 1$ si $(t, uv) \in E$ et $uv_t = 0$ sinon. Quand $v_t = 1$ on dit que le nœud v est impliqué dans S à l’instant t ou que v est présent à l’instant t . Quand $uv_t = 1$ on dit que les nœuds u et v sont liés à l’instant t , ou que le lien uv est présent à l’instant t .

On note T_v l’ensemble des instants auxquels v est présent, et T_{uv} l’ensemble des instants auxquels uv est présent : $T_v = \{t, v_t = 1\}$ et $T_{uv} = \{t, uv_t = 1\}$. On suppose que ces ensembles sont toujours des unions d’intervalles de T fermés et

disjoints (éventuellement réduits à un singleton). Remarquons que $T_{uv} \subseteq T_u \cap T_v$.

De façon duale, on note V_t l’ensemble de nœuds présents à l’instant t , et E_t l’ensemble des liens présents à l’instant t : $V_t = \{v, v_t = 1\}$ et $E_t = \{uv, uv_t = 1\}$.

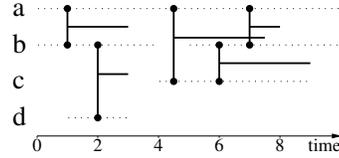


FIGURE 3 – Un *stream graph* $S = (T, V, W, E)$ avec $T = [0, 10] \subseteq \mathbb{R}$, $V = \{a, b, c, d\}$, $W = [0, 10] \times \{a\} \cup ([0, 4] \cup [5, 10]) \times \{b\} \cup [4, 9] \times \{c\} \cup [1, 3] \times \{d\}$, et $E = ([1, 3] \cup [7, 8]) \times \{ab\} \cup [4.5, 7.5] \times \{ac\} \cup [6, 9] \times \{bc\} \cup [2, 3] \times \{bd\}$. En d’autres termes, $T_a = [0, 10]$, $T_b = [0, 4] \cup [5, 10]$, $T_c = [4, 9]$, $T_d = [1, 3]$, $T_{ab} = [1, 3] \cup [7, 8]$, $T_{ac} = [4.5, 7.5]$, $T_{bc} = [6, 9]$, $T_{bd} = [2, 3]$, et $T_{ad} = T_{cd} = \emptyset$.

Si tous les nœuds sont présents tout le temps, *i.e.* $T_v = T$ pour tout v ou, de façon équivalente, $V_t = V$ pour tout t , alors on dit que S est un **flot de liens** et on le note $L = (T, V, E)$ (avec $W = T \times V$ implicitement). En effet, il n’y a pas de dynamique des nœuds dans ce cas, et S est entièrement défini par ce triplet.

Remarquons enfin qu’on considère ici un temps continu représenté par un intervalle de \mathbb{R} . On pourrait faire d’autres choix, comme par exemple un temps discret représenté par un intervalle de \mathbb{N} , comme expliqué dans la section 18 de [3]. Le cas où les ensembles T_{uv} sont des ensembles finis de singletons (comme dans la Figure 1) est également traité dans [3], section 19.

3 Généralisation des graphes

Un graphe $G = (V, E)$ est défini par un ensemble fini de nœuds V et un ensemble de liens $E \subseteq V \otimes V$: $uv \in E$ signifie que u et v sont liés dans G .

Étant donné un stream graph $S = (T, V, W, E)$, on définit $G_t = (V_t, E_t)$, le **graphe induit par S à l’instant t** . Dans le cas de la Figure 3, par exemple, on obtient $G_2 = (\{a, b, d\}, \{ab, bd\})$.

On définit aussi $G(S) = (\{v, T_v \neq \emptyset\}, \{uv, T_{uv} \neq \emptyset\}) = (\bigcup_{t \in T} V_t, \bigcup_{t \in T} E_t)$ le **graphe induit par S** : ses nœuds sont ceux présent dans S et ils sont liés dans $G(S)$ s’il existe un instant auquel ils sont liés dans S . Dans le cas de la Figure 3, $G(S) = (\{a, b, c, d\}, \{ab, ac, bc, bd\})$.

Dans le cas très particulier où $G_t = G_{t'}$ pour tous t et t' dans T , et donc $G(S) = G_t$ pour tout t , alors il n’y a de dynamique ni sur les nœuds, ni sur les liens : tout nœud et tout lien est soit tout le temps présent, soit jamais. On dit alors que S est un **flot graphe-équivalent** : il est entièrement déterminé par le graphe $G(S)$.

De ce point de vue, **les graphes sont donc des flots particuliers**. Par conséquent, les notions usuelles définies sur les

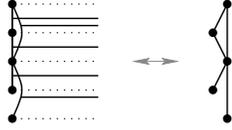


FIGURE 4 – Exemple de flot de liens *graphe-équivalent* (gauche) avec le graphe équivalent (droite).

graphes devraient pouvoir s'étendre au cas des flots, de façon cohérente : si on considère un flot graphe-équivalent, alors la propriété généralisée devrait être la même que sur le graphe équivalent au flot. Nous illustrons cette approche ci-dessous sur deux concepts clés : la densité et les cliques.

3.1 Densité

La densité d'un graphe $G = (V, E)$ est la probabilité, quand on prend une paire de nœuds uv aléatoire dans $V \otimes V$, qu'il y ait un lien entre u et v dans G , *i.e.* $uv \in E$. Autrement dit, c'est la fraction des liens possibles qui existent effectivement : $\delta(G) = |E|/|V \otimes V|$.

On définit la **densité** du stream graph $S = (T, V, W, E)$ comme la probabilité, quand on prend un instant et une paire de nœuds (t, uv) aléatoire dans $T \times V \otimes V$ avec (t, u) et (t, v) dans W , qu'il y ait un lien entre u et v à l'instant t dans S , *i.e.* $(t, uv) \in E$. Autrement dit, c'est la fraction des liens possibles qui existent effectivement :

$$\delta(S) = \frac{\sum_{uv \in V \otimes V} |T_{uv}|}{\sum_{uv \in V \otimes V} |T_u \cap T_v|} = \frac{\int_{t \in T} |E_t| dt}{\int_{t \in T} |V_t \otimes V_t| dt} \quad (1)$$

Dans le cas de la Figure 3, $\sum_{uv \in V \otimes V} |T_{uv}| = |T_{ab}| + |T_{ac}| + |T_{bc}| + |T_{bd}| = 3 + 3 + 3 + 1 = 10$, $\sum_{uv \in V \otimes V} |T_u \cap T_v| = |T_a \cap T_b| + |T_a \cap T_c| + |T_a \cap T_d| + |T_b \cap T_c| + |T_b \cap T_d| + |T_c \cap T_d| = 9 + 5 + 2 + 4 + 2 + 0 = 22$, et on obtient $\delta(S) = \frac{10}{22} \sim 0.45$.

Remarquons que, contrairement aux graphes où $\delta(G) = |E|/|V \otimes V|$, il n'y a pas en général de relation directe entre la densité $\delta(S)$ d'un stream graph et les tailles des ensembles W et E , comme montré Figure 5.

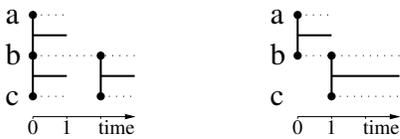


FIGURE 5 – Deux stream graphs avec $|W| = 2 \cdot |T|$ et $|E| = |T|$, mais de densités différentes : à gauche $\delta = 0.75$ et à droite $\delta = 1$.

Toutefois, la relation existe pour les flots de liens (c'est-à-dire si $W = T \times V$). En effet, si on considère $L = (T, V, E)$ on a $T_u = T_v = |T|$ pour tous u et v , et $|W| = |V| \cdot |T|$, ce qui amène :

$$\delta(L) = \frac{\sum_{uv \in V \otimes V} |T_{uv}|}{\sum_{uv \in V \otimes V} |T|} = \frac{2 \cdot |E|}{|V| \cdot (|V| - 1)} \quad (2)$$

De plus, $\delta(L)$ est égale à la densité moyenne des G_t : $\frac{1}{|T|} \int_t \delta(G_t) dt = \frac{1}{|T|} \int_t \frac{|E_t|}{|V_t \otimes V_t|} dt = \frac{1}{|T| \cdot |V \otimes V|} \int_t |E_t| dt = \frac{\int_t |E_t| dt}{\int_t |V_t \otimes V_t| dt} = \delta(L)$, puisque, dans L , $V_t = V$ pour tout t .

Enfin, si S est graphe-équivalent, alors sa densité est égale à la densité de $G(S)$.

3.2 Cliques

Dans un graphe $G = (V, E)$, une clique est un ensemble de nœuds $C \subseteq V$ tous liés deux à deux dans G , *i.e.* $C \otimes C \subseteq E$. De façon équivalente, le sous-graphe induit par C a densité 1.

Dans un stream graph $S = (T, V, W, E)$, on définit une **clique** comme un ensemble de nœuds temporels $I \times X \subseteq W$ tels que les nœuds de X sont tous liés en permanence deux à deux pendant I , *i.e.* $I \times C \otimes C \subseteq E$. De façon équivalente, le sous-flot induit par X pendant I a densité 1.

Dans le cas des graphes comme des flots, une clique est maximale si elle n'est incluse dans aucune autre.

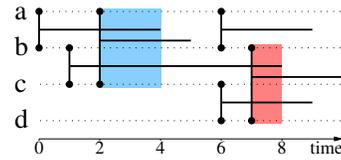


FIGURE 6 – Exemples de cliques maximales. On montre deux cliques maximales impliquant trois nœuds : $[2, 4] \times \{a, b, c\}$ et $[7, 8] \times \{b, c, d\}$.

Par exemple, dans la Figure 6, $[0, 1] \times \{a, b\}$ est une clique. Mais elle n'est pas maximale puis qu'elle est incluse dans $[0, 4] \times \{a, b\}$, qui elle est une clique maximale. Cette clique intersecte une autre clique maximale, $[2, 4] \times \{a, b, c\}$. Il y a une seule autre clique maximale impliquant trois nœuds, $[8, 9] \times \{b, c, d\}$.

Pour tout $[b, e] \subseteq T$ et $X \subseteq V$, si $[b, e] \times X$ est une clique dans S alors X est nécessairement une clique dans $G(S)$. Toutefois, si $[b, e] \times X$ est maximale pour S alors X n'est pas nécessairement maximale pour $G(S)$, comme par exemple $[0, 4] \times \{a, b\}$ dans la Figure 6 ($\{a, b\}$ est une clique dans $G(S)$ mais elle est incluse dans son autre clique $\{a, b, c\}$). Réciproquement, si X est une clique dans $G(S)$ alors en général il n'y a pas de $[b, e]$ tel que $[b, e] \times X$ est une clique dans S .

Si S est graphe-équivalent, alors ses cliques maximales correspondent exactement aux cliques maximales de $G(S)$: $T \times X$ est une clique maximale de S si et seulement si X est une clique maximale de $G(S)$. Par ailleurs, les algorithmes de calcul de cliques maximales des graphes peuvent être étendus au calcul de celles des flots [8].

3.3 Plus de généralisations

Les généralisations de la densité et des cliques présentées ci-dessus illustrent la démarche que nous avons développée dans [3]. Celle-ci consiste d'abord en l'identification des concepts les plus importants de la théorie des graphes (en se basant par

exemple sur les concepts introduits dans la plupart des livres de cours), puis en la recherche des concepts élémentaires sous-jacents, utilisés pour leurs définitions. Une fois ces concepts élémentaires identifiés, il s'agit d'en proposer des généralisations aux flots qui soient aussi simples que possible, et cohérentes entre elles. On peut ensuite définir, à partir de ces concepts élémentaires, les concepts plus évolués qui reposent sur eux. Cette phase permet de confirmer la pertinence des définitions élémentaires, et de les améliorer si besoin. La cohérence avec la théorie des graphes, grâce aux flots graphe-équivalents, ainsi que la cohérence des définitions entre elles, constituent des guides précieux pour ce travail.

Ainsi, nous généralisons dans [3] les concepts élémentaires de théorie des graphes, comme la taille, les sous-graphes, la densité, ou les chemins. Ceci nous permet des généralisations naturelles de concepts plus avancés, comme les voisinages, degrés, coefficient de *clustering*, distances, cliques, composantes connexes, quotients, cœurs, et de nombreux autres. Ces définitions sont cohérentes avec la théorie des graphes, et entre elles. Par exemple, dans les graphes comme dans les flots, le coefficient de clustering d'un nœud est la densité du sous-objet induit par son voisinage. Ces généralisations fournissent un socle pour généraliser des concepts plus avancés comme la modularité, les communautés, ou les marches aléatoires, qui n'existent toujours pas pour les flots de liens.

4 Généraliser les séries temporelles

On peut suivre la même démarche pour généraliser aux flots de liens les notions classiques pour l'analyse de séries temporelles. Il s'agit de montrer que les séries temporelles peuvent être vues comme des cas particuliers de flots de liens, d'identifier les concepts élémentaires utilisés dans l'analyse de séries temporelles, d'en définir les généralisations qui sont cohérentes entre elles, et de s'en servir pour généraliser les notions plus évoluées.

Remarquons d'abord qu'un flot de liens $L = (T, V, E)$ avec seulement deux nœuds u et v , *i.e.* $|V| = 2$, est équivalent à la série $f(t) = 1$ si $(t, uv) \in E$, 0 sinon, voir Figure 7 (gauche). Pour aller plus loin, supposons qu'un ordre sur les paires de nœuds est donné. Un flot très particulier $L = (T, V, E)$ dans lequel les liens apparaissent toujours entre les premières paires de nœuds (dans l'ordre donné) est équivalent à la série $f(t) = |\{uv, (t, uv) \in E\}| = |E_t|$, qu'on appelle l'*activité* de L . On dit alors que L est un **flot série-équivalent**. Voir Figure 7 (droite).

En d'autres termes, une fois l'ordre sur les paires de nœuds donné, l'activité d'un flot série-équivalent code toute l'information du flot. La réciproque est évidente. En ce sens, on peut donc dire que, comme les graphes, **les séries temporelles sont des cas particuliers de flots de liens**.

À partir de ce constat, on peut tenter de définir les concepts clés pour les séries temporelles, comme par exemple la périodicité, la fréquence, la stationnarité, la dérivée, les notions

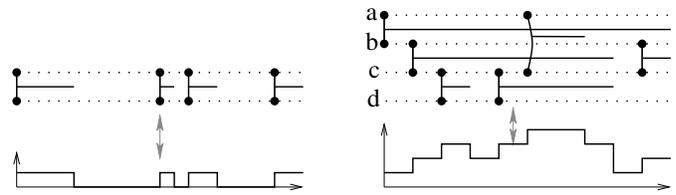


FIGURE 7 – Exemples de flots de liens signal-équivalents. Pour le flot de droite, l'ordre sur les paires de nœuds est ab, bc, cd, ac, bd, ad .

de bruit, de filtre, ou de transformée de Fourier, par exemple. Comme pour les graphes, ces concepts sont ceux introduits dans la plupart des livres de cours sur ce sujet. La littérature du domaine de signal sur graphes est aussi une ressource précieuse [5, 7].

Les définitions de ces notions font appel à des concepts élémentaires comme la variation de la série entre deux points du temps, sa moyenne, sa limite en un point, et d'autres, qui restent à préciser. Les objets les plus pertinents restent eux-mêmes à identifier : peut-être est-il nécessaire de considérer des liens continûment valués plutôt que discrets, comme ci-dessus ? d'utiliser des seuillages ? des fenêtres de temps de tailles variables ? peut être l'activité n'est-elle pas la bonne série temporelle à considérer ? ...

Références

- [1] Batagelj and Praprotnik. An algebraic approach to temporal network analysis based on temporal quantities. *Social Network Analysis and Mining*, 2016.
- [2] Casteigts, Flocchini, Quattrociocchi, and Santoro. Time-varying graphs and dynamic networks. *International Journal of Parallel, Emergent and Distributed Systems*, 2012.
- [3] Matthieu Latapy, Tiphaine Viard, and Clémence Magnien. Stream graphs and link streams for the modeling of interactions over time. *Social Netw. Analys. Mining*, 8(1) :61 :1–61 :29, 2018.
- [4] Léo, Crespelle, and Fleury. Non-altering time scales for aggregation of dynamic networks into series of graphs. In *Proceedings of ACM CoNEXT*, 2015.
- [5] Ortega, Frossard, Kovacevic, Moura, and Vandergheynst. Graph signal processing : Overview, challenges, and applications. *Proceedings of the IEEE*, 2018.
- [6] Ribeiro, Perra, and Baronchelli. Quantifying the effect of temporal resolution on time-varying networks. *Scientific Reports*, 2013.
- [7] Tremblay, Gonçalves, and Borgnat. Design of graph filters and filterbanks. In *Cooperative and Graph Signal Processing*. Academic Press, 2018.
- [8] Viard, Magnien, and Latapy. Enumerating maximal cliques in link streams with durations. *Information Processing Letters*, 2018.