

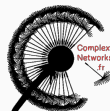
Énumération efficace des cliques maximales dans les flots de liens réels massifs

Alexis BAUDIN*, Clémence MAGNIEN et Lionel TABOURIER

Jeudi 19 janvier 2023

EGC 2023

23ème conférence francophone sur l'extraction et la gestion de connaissances

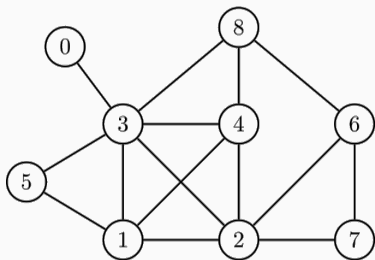


1 - Les cliques dans un graphe

1 - Les cliques dans un graphe

> Définition

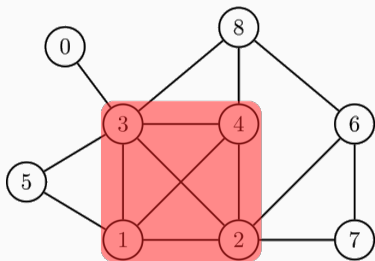
Graphe → interactions statiques



1 - Les cliques dans un graphe

> Définition

Graphe \rightarrow interactions statiques



- $\{1, 2, 3, 4\}$ clique

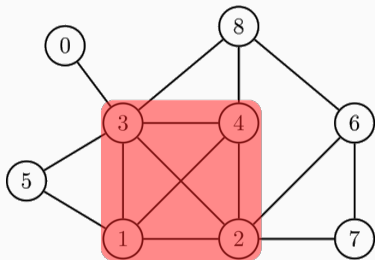
Clique d'un graphe

Ensemble de sommets tous connectés entre eux.

1 - Les cliques dans un graphe

> Définition

Graphe → interactions statiques



- $\{1, 2, 3, 4\}$ clique **maximale**
- $\{1, 2, 3\}$ clique **non maximale**

Clique d'un graphe

Ensemble de sommets tous connectés entre eux.

Clique maximale

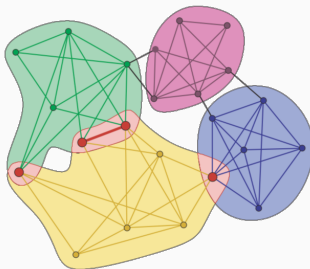
Incluse dans aucune autre clique.

→ Énumération :
Bron-Kerbosch (1973)

1 - Les cliques dans un graphe

> Enjeux

Détection de communautés

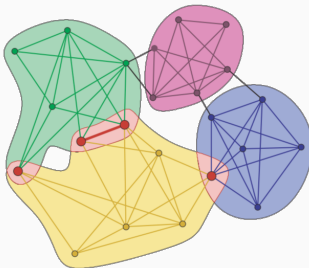


Palla et al. 2005

1 - Les cliques dans un graphe

> Enjeux

Détection de communautés



Palla et al. 2005

→ Ajout d'une dimension temporelle ?

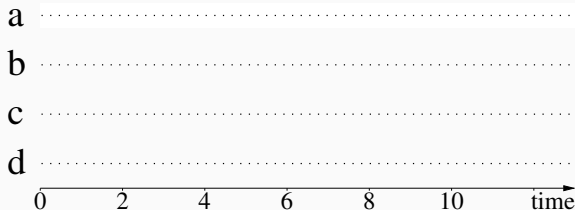
Exemples :

développement d'épidémie, détection d'anomalies (spam, ...)

2 - Flots de liens : données d'interactions temporelles

2 - Flots de liens : données d'interactions temporelles

> Définition

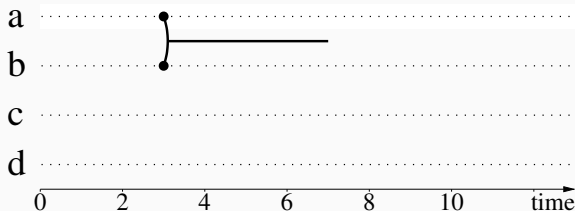


Flot de liens

- 4 sommets : *a*, *b*, *c* et *d*
- Durée : $[0,12]$

2 - Flots de liens : données d'interactions temporelles

> Définition

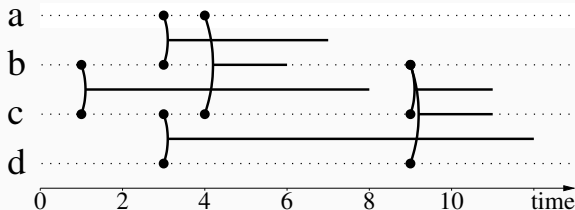


Flot de liens

- 4 sommets : *a*, *b*, *c* et *d*
- Durée : $[0,12]$
- Exemple : *a* et *b* sont connectés sur $[3,7]$

2 - Flots de liens : données d'interactions temporelles

> Définition

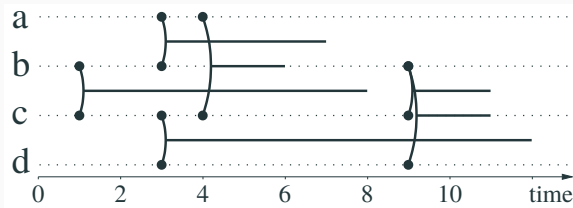


Flot de liens

- 4 sommets : *a*, *b*, *c* et *d*
- Durée : $[0,12]$
- Exemple : *a* et *b* sont connectés sur $[3,7]$

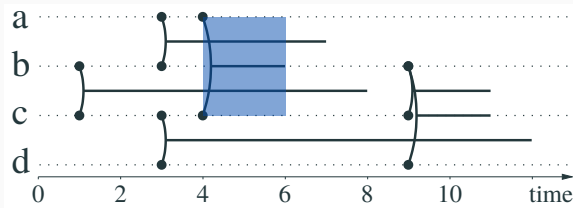
2 - Flots de liens : données d'interactions temporelles

> Cliques maximales dans un flot de liens



2 - Flots de liens : données d'interactions temporelles

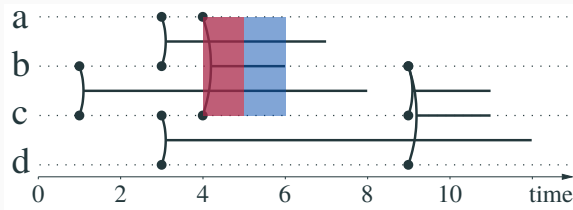
> Cliques maximales dans un flot de liens



$(\{a, b, c\}, [4, 6])$ est une clique

2 - Flots de liens : données d'interactions temporelles

> Cliques maximales dans un flot de liens

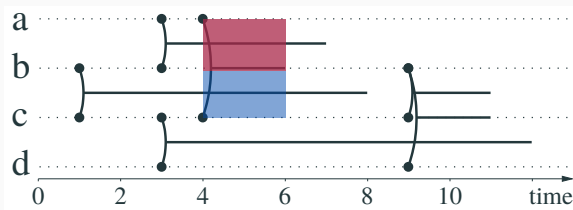


$(\{a, b, c\}, [4, 6])$ est une clique

→ $(\{a, b, c\}, \underline{[4, 5]})$ n'est pas maximale en temps

2 - Flots de liens : données d'interactions temporelles

> Cliques maximales dans un flot de liens



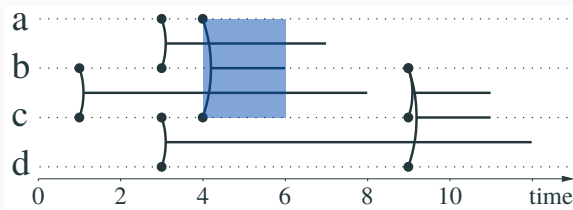
$(\{a, b, c\}, [4, 6])$ est une clique

→ $(\{a, b, c\}, [4, 5])$ n'est pas maximale en temps

→ $(\{a, b\}, [4, 6])$ n'est pas maximale en sommets

2 - Flots de liens : données d'interactions temporelles

> Cliques maximales dans un flot de liens



$(\{a, b, c\}, [4, 6])$ est une clique

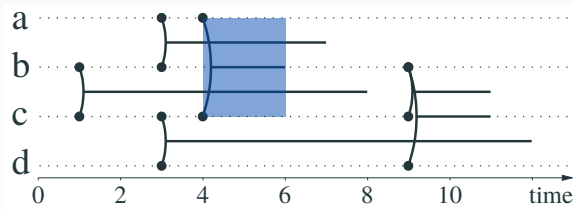
→ $(\{a, b, c\}, [4, 5])$ n'est pas maximale en temps

→ $(\{a, b\}, [4, 6])$ n'est pas maximale en sommets

maximale = en temps et en sommets

2 - Flots de liens : données d'interactions temporelles

> Cliques maximales dans un flot de liens



$(\{a, b, c\}, [4, 6])$ est une clique maximale.

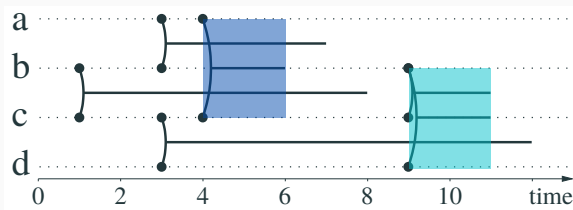
→ $(\{a, b, c\}, [4, 5])$ n'est pas maximale en temps

→ $(\{a, b\}, [4, 6])$ n'est pas maximale en sommets

maximale = en temps et en sommets

2 - Flots de liens : données d'interactions temporelles

> Cliques maximales dans un flot de liens



$(\{a, b, c\}, [4, 6])$ est une clique maximale.

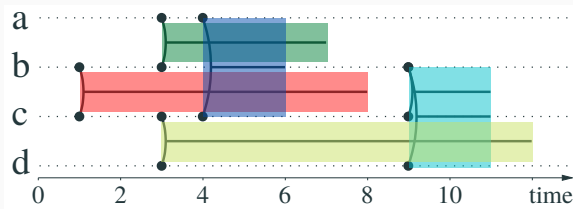
→ $(\{a, b, c\}, [4, 5])$ n'est pas maximale en temps

→ $(\{a, b\}, [4, 6])$ n'est pas maximale en sommets

maximale = en temps et en sommets

2 - Flots de liens : données d'interactions temporelles

> Cliques maximales dans un flot de liens



$(\{a, b, c\}, [4, 6])$ est une clique maximale.

→ $(\{a, b, c\}, [4, 5])$ n'est pas maximale en temps

→ $(\{a, b\}, [4, 6])$ n'est pas maximale en sommets

maximale = en temps et en sommets

3 - Énumération des cliques maximales dans les flots de liens

3 - Énumération des cliques maximales dans les flots de liens

> État de l'art

État de l'art : quatre travaux principaux

- Viard *et al.* 2016
- Viard *et al.* 2018
- Himmel *et al.* 2017
- Bentert *et al.* 2019

3 - Énumération des cliques maximales dans les flots de liens

> État de l'art

État de l'art : quatre travaux principaux

- Viard *et al.* 2016
 - Viard *et al.* 2018
 - Himmel *et al.* 2017
 - Bentert *et al.* 2019
- } Mémoïsation
⇒ trop de mémoire
- } Besoin de toutes les interactions
passées et futures

⇒ **Problème de passage à l'échelle.**

3 - Énumération des cliques maximales dans les flots de liens

> État de l'art

État de l'art : quatre travaux principaux

- Viard *et al.* 2016
 - Viard *et al.* 2018
 - Himmel *et al.* 2017
 - Bentert *et al.* 2019
- } Mémoïsation
⇒ trop de mémoire
- } Besoin de toutes les interactions
passées et futures

⇒ **Problème de passage à l'échelle.**

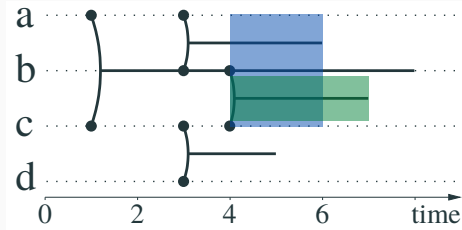
Nouvel algorithme :

- cliques non conservées en mémoire ;
- interactions réduites à chaque pas de temps.

3 - Énumération des cliques maximales dans les flots de liens

> Algorithme

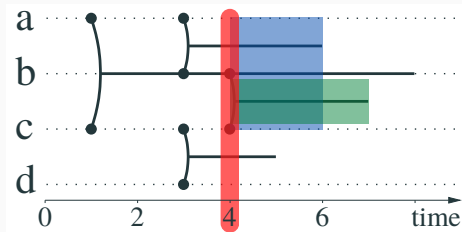
Cliques maximales qui commencent à $t = 4$



3 - Énumération des cliques maximales dans les flots de liens

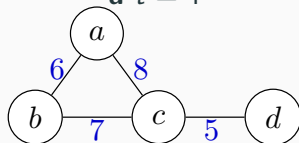
> Algorithme

Cliques maximales qui commencent à $t = 4$



Graphe instantané G_t

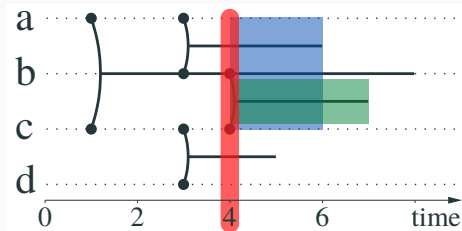
à $t = 4$



3 - Énumération des cliques maximales dans les flots de liens

> Algorithme

Cliques maximales qui commencent à $t = 4$

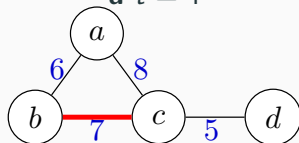


$(\{b, c\}, \quad), (\{b, c, a\}, \quad)$

→ cliques de G_t contenant une nouvelle arête ;

Graphe instantané G_t

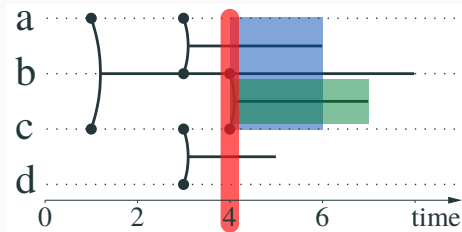
à $t = 4$



3 - Énumération des cliques maximales dans les flots de liens

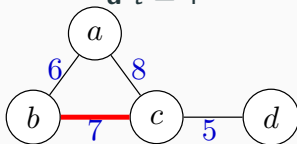
> Algorithme

Cliques maximales qui commencent à $t = 4$



Graphe instantané G_t

à $t = 4$



$(\{b, c\}, [4, \infty))$, $(\{b, c, a\}, [4, \infty))$

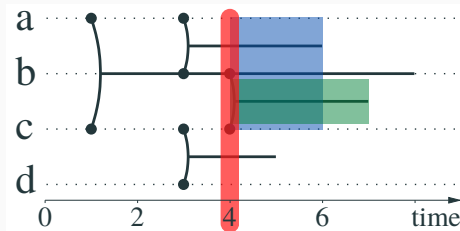
→ cliques de G_t contenant une nouvelle arête ;

→ temps début : $t = 4$;

3 - Énumération des cliques maximales dans les flots de liens

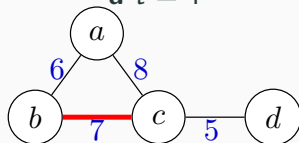
> Algorithme

Cliques maximales qui commencent à $t = 4$



Graphe instantané G_t

à $t = 4$



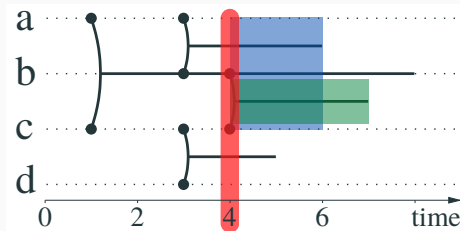
$(\{b, c\}, [4, 7]), (\{b, c, a\}, [4, 6])$

- cliques de G_t contenant une nouvelle arête ;
- temps début : $t = 4$;
- temps fin : min des temps de fin des arêtes.

3 - Énumération des cliques maximales dans les flots de liens

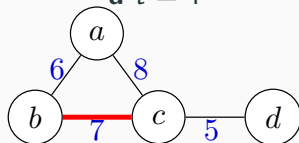
> Algorithme

Cliques maximales qui commencent à $t = 4$



Graphe instantané G_t

à $t = 4$



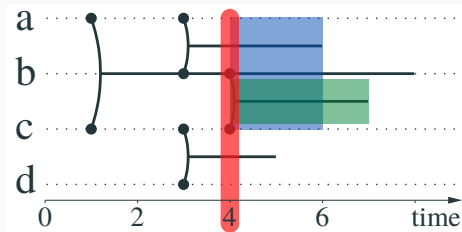
$((\{b, c\}, [4, 7]), (\{b, c, a\}, [4, 6]))$

- cliques de G_t contenant une nouvelle arête ;
- temps début : $t = 4$;
- temps fin : min des temps de fin des arêtes.

3 - Énumération des cliques maximales dans les flots de liens

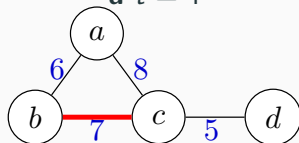
> Algorithme

Cliques maximales qui commencent à $t = 4$



Graphe instantané G_t

à $t = 4$



$((b, c), [4, 7]), ((b, c, a), [4, 6])$

- cliques de G_t contenant une nouvelle arête ;
- temps début : $t = 4$;
- temps fin : min des temps de fin des arêtes.

⇒ On se ramène à un problème de graphes : Bron-Kerbosch.

3 - Énumération des cliques maximales dans les flots de liens

> Algorithme

Bilan du nouvel algorithme

- cliques non conservées en mémoire ✓
- interactions réduites à chaque pas de temps ✓

4 - Étude expérimentale : gain de performance

4 - Étude expérimentale : gain de performance

Protocole expérimental

- Implémentations Python et C++
- Maximum 24h et 390Gb de RAM.

4 - Étude expérimentale : gain de performance

Protocole expérimental

- Implémentations Python et C++
- Maximum 24h et 390Gb de RAM.

Jeux de données : état de l'art + flots de liens massifs

- réseaux de communication
- interactions humaines
- interactions biologiques

4 - Étude expérimentale : gain de performance

Flot de liens	#liens	#cliques
highschool-2011	6 472	7 732
facebooklike	50 056	50 080
infectious	100 329	138 670
stackexchange	870 128	894.317
youtube	$12 \cdot 10^6$	$12 \cdot 10^6$
wikipedia	$39 \cdot 10^6$	$41 \cdot 10^6$
soc-bitcoin	$94 \cdot 10^6$	$787 \cdot 10^6$

4 - Étude expérimentale : gain de performance

Flot de liens	#liens	#cliques	Litt.	Python	C++
highschool-2011	6 472	7 732	1.2s	0.12s	0.05s
facebooklike	50 056	50 080	15s	0.42s	0.12s
infectious	100 329	138 670	20s	1.7s	0.78s
stackexchange	870 128	894.317	1h	10s	1.9s
youtube	$12 \cdot 10^6$	$12 \cdot 10^6$	×	9min	2min
wikipedia	$39 \cdot 10^6$	$41 \cdot 10^6$	×	4h30	1min20s
soc-bitcoin	$94 \cdot 10^6$	$787 \cdot 10^6$	×	×	7h30

Conclusion

Contributions

- nouvel algorithme ;
- gain de 2 ordres de grandeur de taille de données ;
- protocole expérimental sur données réelles ;
- étude détaillée de la complexité ;
- version parallèle.

Contributions

- nouvel algorithme ;
- gain de 2 ordres de grandeur de taille de données ;
- protocole expérimental sur données réelles ;
- étude détaillée de la complexité ;
- version parallèle.

Perspectives – flots de liens

- communautés par percolation de cliques.
- adapter l'algorithme à la recherche de motifs ;

Merci pour votre attention !

Accès au code :

`https://gitlab.lip6.fr/audin/maxcliques-linkstream`

Contact : `alexis.audin@lip6.fr`