

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

Alexis Baudin*, Clémence Magnien*, Lionel Tabourier*

* Sorbonne Université, CNRS, LIP6, F-75005 Paris, France
prenom.nom@lip6.fr

Résumé. Les flots de liens offrent un formalisme de description d’interactions au cours du temps. Un lien correspond à deux sommets qui interagissent sur un intervalle de temps. Une clique est un ensemble de sommets associé à un intervalle de temps durant lequel ils sont tous connectés. Elle est maximale si ni son ensemble de sommets ni son intervalle de temps ne peuvent être augmentés. Les algorithmes existants pour énumérer ces structures ne permettent pas de traiter des jeux de données réels de plus de quelques centaines de milliers d’interactions. Or, l’accès à des données toujours plus massives demande d’adapter les outils à de plus grandes échelles. Nous proposons alors un algorithme qui énumère les cliques maximales sur des réseaux temporels réels et massifs atteignant jusqu’à plus de 100 millions de liens. Nous montrons expérimentalement qu’il améliore l’état de l’art de plusieurs ordres de grandeur.

1 Introduction

L’analyse des réseaux d’interactions issus du monde réel a récemment fait des progrès significatifs, en passant de représentations statiques à des représentations dynamiques. La disponibilité de données temporelles, ainsi que le développement d’outils pour les décrire et les analyser, ont révélé l’importance de la temporalité des événements pour comprendre la structure et le fonctionnement de systèmes complexes en interaction, tels que les réseaux de machines, les réseaux électriques, les réseaux sociaux en ligne, ou même les réseaux biologiques.

Dans cet article, pour représenter les interactions dynamiques, nous utilisons le formalisme des flots de liens, introduit par Latapy et al. (2018). Il associe à chaque interaction un intervalle de temps, pendant lequel l’interaction existe dans le réseau. C’est un formalisme rigoureux qui permet de rendre compte à la fois des aspects temporels et structurels des données d’interaction, ce qui enrichit leur analyse, en considérant ces deux aspects intrinsèquement liés.

Le problème de l’énumération des cliques maximales dans un graphe est NP-difficile. Néanmoins, c’est un problème important pour l’analyse des graphes issus de données d’interactions réelles, car il permet de décrire leur structure. Par exemple, l’énumération de cliques est utilisée pour détecter des sous-graphes denses pertinents (Gibson et al., 2005; Fratkin et al., 2006), pour définir des communautés (Palla et al., 2005; Baudin et al., 2022), ou encore pour la compression de graphes (Buehrer et Chellapilla, 2008).

En y ajoutant une dimension temporelle, l’énumération des cliques maximales dans les flots de liens (maximales à la fois en nombre de sommets et en temps) a suscité de l’intérêt