Génération des cliques maximales d'un graphe, des bicliques maximales d'un graphe biparti

A. Gély L. Nourine
Université Blaise Pascal, Clermont-Ferrand II, France

B. Sadi
Université de Tizi-Ouzou, Algérie

{gely,nourine}@isima.fr, sadibach@yahoo.fr

Résumé

Nous présentons un cadre général permettant d'étudier les algorithmes de génération des cliques maximales d'un graphe et des bicliques maximales d'un graphe biparti. Soit un graphe G, nous définissons un graphe de transitions $T(G)$ dont les sommets sont les cliques maximales du graphe G et les arcs des transitions possibles entre cliques. Nous montrons que $T(G)$ est fortement connexe. Nous caractérisons les arbres de couvertures de ce graphe qui apparaissent implicitement dans plusieurs algorithmes existants [3, 6].

Lorsque G est un graphe biparti, nous montrons que le treillis de Galois de G est un graphe partiel de $T(G)$ et nous en déduisons que les algorithmes de générations des bicliques maximales d'un graphe biparti sont des parcours particulier de $T(G)$. Cela permet de faciliter les comparaisons entre ces algorithmes. Cette comparaison nous a permis de mettre en avant les similitudes de deux algorithmes [1, 5], affinant l'étude de complexité de [1]. Enfin, nous étudions les propriétés du graphe de transitions $T(G)$ pour certaines classes de graphes G. Ainsi, pour une numérotation particulière des sommets, nous montrons qu'il existe un chemin hamiltonien dans $T(G)$ lorsque G est un graphe triangulé ou un graphe de comparabilité.

Mots-clés : Cliques maximales, bicliques maximales, treillis, algorithmes de génération.

1 Introduction

Nous nous intéressons à la génération de toutes les cliques maximales d'un graphe, ainsi qu'à la génération de toutes les bicliques maximales d'un graphe biparti. Ces structures sont utilisées dans de nombreux domaines, notamment lorsqu'il s'agit de faire ressortir des points communs dans des données. La bio-informatique utilise par exemple la génération de cliques maximales pour rechercher des similitudes entre gènes. L'extraction de connaissances (Datamining) fait un usage intensif de la famille des bicliques maximales d'un graphe biparti.

Dans [4], les auteurs considèrent les algorithmes [3, 6] comme des exemples de la technique de recherche inversée[1] initiée dans [2]. Dans leur papier, les auteurs décrivent comment utiliser du calcul matriciel pour améliorer l'efficacité pratique des algorithmes de génération des cliques maximales.

Ici, le but diffère : Nous présentons un cadre général dédié à l'étude des algorithmes de génération des cliques maximales d'un graphe, et des bicliques maximales d'un graphe biparti.

Soit un graphe G, on notera $T(G)$ le graphe dont les sommets sont les cliques maximales de G et les arcs des transitions possibles entre les cliques. Pour une fonction de transition que nous détaillerons par la suite, nous montrerons que $T(G)$ est fortement connexe. D'autre part, nous donnerons les arbres de couvertures de $T(G)$ qui correspondent aux algorithmes [3, 6].

Ayant défini $T(G)$, le graphe de transitions des cliques maximales d'un graphe, nous nous intéresserons à ses propriétés lorsque G appartient à certaines classes. Ainsi, pour une numérotation particulière des sommets, nous montrerons que $T(G)$ possède un chemin hamiltonien lorsque G est un graphe triangulé ou un graphe de comparabilité.

Enfin, nous transposerons ces résultats à la génération des bicliques maximales d'un graphe biparti.

Définitions

On note $G = (V, E)$ un graphe, avec V l'ensemble de ses sommets, E l'ensemble de ses arêtes. $|V| = n$, $|E| = m$. Un graphe biparti G, noté $G = (U, V, E)$ est un graphe $(U \cup V, E)$ tel que U et V sont des ensembles stables.

Le voisinage d'un sommet v sera noté $\Gamma(v) = \{v' \in V \mid vv' \in E\}$.

Définition 1 (Clique) *Une clique* (resp. stable) *de $G = (V, E)$ est un ensemble de sommet $C \subseteq V$, tel que pour tout v_1 et v_2 appartenant à C, l'arête v_1v_2 appartient* (resp. n'appartient pas) *à E.*

Une clique C de $G = (V, E)$ est maximale si pour tout $x \in V \setminus C$, $C \cup \{x\}$ n'est pas une clique. L'ensemble des cliques maximales de G sera noté $\mathcal{C}(G)$.

Graphe de transitions

Nous définissons le graphe de transitions $T(G)$ des cliques maximales d'un graphe G comme le graphe orienté ayant pour ensemble de sommets l'ensemble des cliques maximales de G. Pour $C, C' \in \mathcal{C}(G)$, (C, C') appartient à l'ensemble

[1] reverse search technique, traduction libre

des arcs de $T(G)$ si on peut passer de C à C' en temps polynomial en utilisant une transformation préalablement fixée.

Si $T(G)$ est fortement connexe (ce qui est le cas), alors il est possible de considérer n'importe quel parcours de ce graphe comme une manière d'énumérer les cliques maximales de G. Par suite, il est intéressant de rechercher des arborescences recouvrantes enracinées de $T(G)$ vérifiant certaines propriétés, ou bien de rechercher des chemins hamiltoniens dans $T(G)$.

Soit $G = (V, E)$ avec $V = \{1, \ldots, n\}$. On note par G_i le sous graphe de G induit par les sommets $\{1, \ldots, i\}$. Pour une clique maximale $C \in \mathcal{C}(G)$ et un sommet $i \notin C$, on note $K_i(C)$ l'ensemble de sommets correspondant à la plus petite (au sens lexicographique) clique maximale du graphe induit par les sommets $((C \cap \{1, \ldots, i\}) \cap \Gamma(i)) \cup \{i, \ldots, n\}$.

$K_i(C)$ existe toujours, et il y a deux possibilités :

- $K_i(C)$ est une clique *maximale* de G.
- $K_i(C)$ n'est pas une clique *maximale* de G. Dans ce cas, les sommets manquants pour que $K_i(C)$ soit maximale sont des sommets du graphe G_i.

Pour $C, C' \in \mathcal{C}(G)$, on dira qu'il y a transition possible de C' vers C s'il existe un sommet $i \notin C'$ tel que $K_i(C') = C$. Autrement dit, il existe une transition de C' à C si $K_i(C')$ est la plus petite (au sens lexicographique) clique maximale de G ayant $((C' \cap \{1, \ldots, i\}) \cap \Gamma(i)) \cup \{i\}$ pour préfixe.

Définition 2 (Graphe de Transitions) *Le graphe de transitions des cliques maximales d'un graphe G, noté $T(G) = (\mathcal{C}(G), T)$ est le graphe défini par :*

- *son ensemble de sommets : $\mathcal{C}(G)$, l'ensemble des cliques maximales de G*
- *son ensemble d'arêtes : (C', C) appartient à T ssi il existe une transition de C' vers C.*
- *L'arc (C', C) est étiqueté avec le sommet i permettant la transition de C' vers C, noté $Label(C', C) = i$.*

A partir de cette définition, nous présenterons quelques propriétés de ce graphe de transitions pour l'énumération des cliques maximales d'un graphe et pour l'énumération des bicliques maximales d'un graphe biparti.

Références

[1] G. Alexe, S. Alexe, Y. Crama, S. Foldes, P. L. Hammer, and B. Simeone. Consensus algorithms for the generation of all maximal bicliques. *Discrete Applied Mathematics*, 145(1) :11–21, december 2004.

[2] D. Avis and K. Fukuda. Reverse search for enumeration. *Discrete Applied Mathematics*, 65 :21–46, 1996.

[3] D. S. Johnson, M. Yannakakis, and C. H. Papadimitriou. On generating all maximal independent sets. *Information Processing Letters*, 27 :119–123, March 1988.

[4] K. Makino and T. Uno. New algorithms for enumerating all maximal cliques. In *Proc. 9th Scand. Worksh. Algorithm Theory (SWAT 2004)*, LNCS, pages 260–272, 2004.

[5] L. Nourine and O. Raynaud. A fast algorithm for building lattices. *Information Processing Letters*, 71 :199–204, 1999.

[6] S. Tsukiyama, M. Ide, M. Aiyoshi, and I. Shirawaka. A new algorithm for generating all the independent sets. *SIAM J. Computing*, 6 :505–517, december 1977.