IT-3004 - Graphes et algorithmes - TD 1 - Michel Couprie

Cyber-Profiler

Dans un logiciel d'aide à la résolution d'affaires criminelles, une structure de graphe non orienté est utilisée pour rechercher des corrélations entre différentes affaires. Un sommet du graphe représente un évènement, une personne ou un lieu, et deux sommets sont reliés par une arête s'il est possible d'établir un lien entre les deux, par exemple, si une personne p a séjourné dans un lieu ℓ, ou a participé à un évènement e, ou encore si deux personnes p_1 et p_2 ont un lien familial ou professionnel...

Le logiciel est utilisé pour rechercher automatiquement des groupes de sommets qui sont liés mutuellement par de nombreuses arêtes, ces coïncidences pouvant indiquer des liens entre plusieurs enquêtes présentes ou passées. Dans ce qui suit, on cherchera à formaliser et à quantifier cette notion de "coïncidence", ainsi qu'à trouver une méthode pour les détecter.

Soit $G = (E, \vec{\Gamma})$ un graphe et soit X une partie de E. On appelle *sous-graphe de G induit par X*, et l'on note G_X, le graphe défini par $G_X = (X, \vec{\Gamma} \cap [X \times X])$. Autrement dit, tous les arcs de G dont les deux extrémités sont dans X, et seulement ceux-ci, se retrouvent dans G_X.

On dit que deux sommets distincts x *et* y *sont adjacents dans* G s'ils forment une arête de G, autrement dit si (x, y) ou (y, x) est un arc de G.

Dans la suite, on supposera que le graphe $G = (E, \vec{\Gamma})$ est symétrique et sans boucle (anti-réflexif). On notera respectivement n et m le nombre de sommets et d'arcs de G ($n = |E|$, $m = |\vec{\Gamma}|$).

Soit X une partie de E, et soit $k \in \mathbb{N}$. On dit que X *est k-lié dans* G si chaque sommet $x \in X$ est adjacent à au moins k autres sommets dans le sous-graphe de G induit par X.

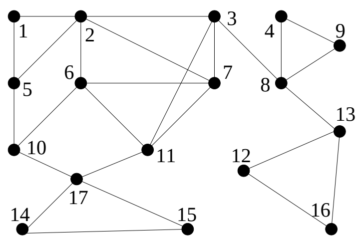

Figure 1

a) Dans le graphe de la figure 1, peut-on trouver :
— un ensemble 1-lié (mais non 2-lié) comportant 3 sommets ?
— un ensemble 2-lié (mais non 3-lié) comportant 4 sommets ?
— un ensemble 3-lié comportant 5 sommets ?
— un ensemble 4-lié ?
Dans chaque cas de réponse positive, donnez l'ensemble que vous avez trouvé.

b) Soit $x \in E$. On note $d(x, G)$ le *degré de x dans G*, c'est-à-dire le nombre de sommets de E qui sont adjacents à x dans G. Reformulez de manière compacte la définition d'un ensemble k-lié en utilisant cette notation.

c) Soit $X \subseteq E$, que pensez-vous de l'affirmation suivante :
X est 1-lié dans G si et seulement si X est connexe.
Justifiez votre réponse.

Dans notre application à la recherche de liens entre affaires criminelles, on recherchera les ensembles k-liés (non vides) où k est le plus grand possible, le nombre k étant un indice de la richesse d'un faisceau de coïncidences. Mais cela suffit-il toujours à montrer la connexion entre deux affaires différentes ? La question suivante apporte un élément de réponse à cette question.

d) Soit $X \subseteq E$ avec $|X| = p \geq 2$. Montrez que si X est k-lié avec $k \geq p/2$ alors X est connexe. Cela revient à montrer que tout graphe à n sommets dont chaque sommet a un degré supérieur ou égal à $n/2$ est connexe.

e) Soit $k \in \mathbb{N}$. Proposez une méthode pour trouver automatiquement, s'il en existe, un sous-ensemble X de E qui soit k-lié dans G. Votre méthode devra, en cas de non-existence d'un tel X, se terminer en donnant une réponse appropriée. Vous pouvez décrire votre méthode en français, une question ultérieure demandera un schéma de programme.

f) Proposer un schéma de programme détaillé pour la méthode de la question e). Il n'est pas demandé, pour le moment, de trouver un algorithme de complexité minimale.

g) Évaluez la complexité de cet algorithme.

h) Si votre algorithme (questions f-g) n'est pas linéaire, proposez un autre algorithme en $O(n + m)$. Justifiez sa complexité.

i) Exécutez l'algorithme "à la main" sur le graphe de la Fig. 1 pour $k = 3$.

On va maintenant chercher des arguments pour justifier la méthode proposée, en d'autres termes, prouver qu'elle donne toujours le résultat attendu (conforme à la définition).

Soit X un sous-ensemble de E. On dit que X est une *clique de G* si tout sommet de X est adjacent à tous les autres sommets de X. On dit que la clique X est *maximale* (au sens de l'inclusion) s'il n'existe pas de clique Y de G qui contienne strictement X.

j) Quel est le lien entre clique et ensemble k-lié ? Illustrer la notion de clique maximale par quelques exemples. Donner un exemple de graphe dans lequel on trouve deux cliques maximales différentes mais ayant une intersection non vide.

k) Soit X un sous-ensemble k-lié de E maximal pour l'inclusion. Montrer que ce sous-ensemble est unique.

l) Soit X un sous-ensemble de E et soit Y un sous-ensemble de E qui inclut X. Montrer que X est k-lié pour G si et seulement si X est k-lié pour le sous-graphe induit G_Y.

m) On note E_k le sous-ensemble k-lié de E maximal au sens de l'inclusion (voir question k). Montrer que le résultat de la méthode trouvée en e) est précisément E_k.

n) Montrer que si $k > k'$, alors $E_k \subseteq E_{k'}$.

o) Proposer un algorithme *linéaire en temps de calcul* pour trouver automatiquement le plus grand entier k tel qu'il existe un sous-ensemble X non vide de E qui soit k-lié dans G. On s'appuiera sur la propriété n) pour éviter les calculs redondants.