Isomorphisme et similarités de graphes

Épreuve pratique d'algorithmique et de programmation
Concours commun des Écoles normales supérieures

Durée de l'épreuve : 3 heures 30 minutes

Juin/Juillet 2022

ATTENTION !

N'oubliez en aucun cas de recopier votre u_0
à l'emplacement prévu sur votre fiche réponse

Important.

Il vous a été donné un numéro u_0 qui servira d'entrée à vos programmes. Les réponses attendues sont généralement courtes et doivent être données sur la fiche réponse fournie à la fin du sujet. À la fin du sujet, vous trouverez en fait deux fiches réponses. La première est un exemple des réponses attendues pour un $\widetilde{u_0}$ particulier (précisé sur cette même fiche et que nous notons avec un tilde pour éviter toute confusion !). Cette fiche est destinée à vous aider à vérifier le résultat de vos programmes en les testant avec $\widetilde{u_0}$ au lieu de u_0. Vous indiquerez vos réponses (correspondant à votre u_0) sur la seconde et vous la remettrez à l'examinateur à la fin de l'épreuve.

En ce qui concerne la partie orale de l'examen, lorsque la description d'un algorithme est demandée, vous devez présenter son fonctionnement de façon schématique, courte et précise. Vous ne devez en aucun cas recopier le code de vos procédures !
Quand on demande la complexité en temps ou en mémoire d'un algorithme en fonction d'un paramètre n, on demande l'ordre de grandeur en fonction du paramètre, par exemple : $O(n^2)$, $O(n \log n)$,...
Il est recommandé de commencer par lancer vos programmes sur de petites valeurs des paramètres et de ***tester vos programmes sur des petits exemples que vous aurez résolus préalablement à la main ou bien à l'aide de la fiche réponse type fournie en annexe***. Enfin, il est recommandé de lire l'intégralité du sujet avant de commencer afin d'effectuer les bons choix de structures de données dès le début.

1 Préliminaires

Un graphe non orienté$G = (V, E)$ est défini par la donnée de son ensemble $V \subseteq \mathbb{N}$ de sommets et d'un ensemble E de paires de sommets. On dit qu'il existe une arête entre u et v si $\{u, v\} \in E$. Tout au long de ce sujet, on considérera qu'un sommet ne peut pas avoir un noeud vers lui-même : $\{u, u\} \in E$ est impossible. Par simplicité, on ne considérera dans ce sujet que des graphes dont l'ensemble de sommets est de la forme $\{0, \ldots, n-1\}$, où n est le nombre de sommets du graphe. Par abus de notation, on écrira donc de façon équivalente $\{u, v\} \in G$ pour signifier que $\{u, v\}$ fait partie de l'ensemble des arêtes de G.

Génération de nombres. Étant donné u_0, on définit par récurrence

$$u_{t+1} := 19\ 999\ 991 u_t \pmod{19\ 999\ 999}$$

pour tout $t \in \mathbb{N}$, c'est-à-dire u_{t+1} est le reste de la division euclidienne par 19 999 999 du produit entre 19 999 991 et u_t.

Question 1 *Calculer les valeurs de u_t pour les valeurs de paramètres suivantes :*

- ***a)*** $u_{50} \pmod{1000}$
- ***b)*** $u_{100} \pmod{1000}$
- ***c)*** $u_{1\ 000} \pmod{1000}$
- ***d)*** $u_{5\ 000} \pmod{1000}$

Génération de graphes. Étant donnés deux paramètres $n, p \in \mathbb{N}$, nous allons générer à partir de la suite $(u_t)_{t\in\mathbb{N}}$ un graphe $G_{n,p}$. Pour cela, nous nous appuyons sur la suite $(v_t)_{t\in\mathbb{N}}$ obtenue à partir de $(u_t)_{t\in\mathbb{N}}$ comme suit :

$$v_t := \begin{cases} 1 & \text{si } u_t \pmod{10\ 000} < p \\ 0 & \text{sinon} \end{cases}$$

pour tout $t \in \mathbb{N}$.
Pour construire le graphe $G_{n,p}$ à n sommets, nous allons utiliser les valeurs de la suite v_t pour décider si une arête donnée appartient au graphe ($v_t = 1$) ou n'y appartient pas ($v_t = 0$). Plus précisément, les $n-1$ premières valeurs de v_t sont caractéristiques des arêtes reliant le sommet 0 aux autres sommets : v_0 vaut 1 si et seulement si $G_{n,p}$ admet une arête entre les sommets 0 et 1, i.e. $\{0, 1\} \in G_{n,p}$, v_1 si $\{0, 2\} \in G_{n,p}$, etc. Les $n-2$ valeurs suivantes sont caractéristiques des arêtes reliant 1 aux autres sommets : $\{1, 2\} \in G_{n,p}$ si et seulement si $v_{n-1} = 1$, etc. Et ainsi de suite jusqu'à $v_{\binom{n}{2}-1} = 1$ si et seulement si $\{n-2, n-1\} \in G_{n,p}$.

Graphes colorés. Nous allons travailler avec des graphes dont les sommets sont colorés par des entiers. Sauf mention explicite, nous associerons dans ce sujet à un graphe à n sommets la fonction de coloriage de graphe $c^n : [\![0, n-1]\!] \to [\![0, n-1]\!]$ obtenue à partir de $(u_t)_{\mathbb{N}}$ comme suit :

$$c^n(i) = u_i \pmod{n}$$

Question 2 *Calculer le nombre d'arêtes et de couleurs distinctes dans $G_{n,p}$ pour les valeurs de paramètres suivantes.*

- ***a)*** $n = 10, p = 500$
- ***b)*** $n = 100, p = 5\ 000$
- ***c)*** $n = 100, p = 50$
- ***d)*** $n = 1\ 000, p = 7\ 500$

Patch. Étant donné un graphe G, nous allons considérer un second graphe G^σ obtenu en appliquant une série σ de modifications à G. Les modifications que l'on considère peuvent être de trois natures. La modification $Swap(n, m)$ échange l'ensemble des voisins des sommets m et n :

— pour tout sommet i tel que $i \neq n$ et $i \neq m$: $\{i, n\} \in G$ si et seulement si $\{i, m\} \in G^\sigma$ et $\{i, m\} \in G$ si et seulement si $\{i, n\} \in G^\sigma$.

— si $\{n, m\} \in G$, on conserve cette arête dans G^σ.

On précise bien qu'un *Swap* n'impacte pas les couleurs des sommets. La modification $Add(n, m)$ ajoute l'arête $\{n, m\}$ à l'ensemble des arêtes (sans effet si l'arête existait déjà, ou si $n = m$). Enfin, la modification $Rem(n, m)$ retire l'arête $\{n, m\}$ à l'ensemble des arêtes (sans effet si l'arête n'était déjà pas présente). Un patch σ est une liste de modifications.
Le graphe G^σ est obtenu en appliquant à G les modifications contenues dans σ en séquence, **en respectant pour ordre d'application** de commencer en tête de la liste, vers sa queue : c'est-à-dire que $G^{[m_1,m_2,\ldots,m_n]}$ est égal à $G_1^{[m_2,\ldots,m_n]}$ où G_1 est le graphe résultant de l'application de m_1 à G.
Étant donné $n, p \in \mathbb{N}$, nous considérons le schéma suivant pour générer une modification.

$$m(n,i,j) := \begin{cases} Swap(u_j \mod n, u_{j+1} \mod n) & \text{si } u_i \pmod 3 = 0 \\ Add(u_j \mod n, u_{j+1} \mod n) & \text{si } u_i \pmod 3 = 1 \\ Rem(u_j \mod n, u_{j+1} \mod n) & \text{si } u_i \pmod 3 = 2 \end{cases}$$

Enfin, nous considérons le patch à k modifications $\sigma(n, k)$ défini comme

$$[m(n, n, 2n), m(n, n+1, 2n+1), \ldots, m(n, n+k-1, 2n+k-1)]$$

et notons $\widehat{G}_{n,p}^k$ le graphe obtenu en appliquant le patch $\sigma(n, k)$ à $G_{n,p}$.

Somme de contrôle d'un graphe Pour un graphe G à n sommets, on définit sa somme de contrôle $SC(G)$ à l'aide de la formule suivante :

$$SC(G) = \left(\sum_{x \in [\![0, n-1]\!]} \sum_{y \mid \{x,y\} \in G} (x+y)^2 + y \right) \mod 1\,000\,000$$

Question 3 *Calculer la somme de contrôle des graphes $G_{n,p}$ et $\widehat{G}_{n,p}^k$ pour les valeurs de paramètres suivantes :*

a) $n = 10, p = 500, k = 10$

b) $n = 100, p = 5\ 000, k = 10$

c) $n = 100, p = 50, k = 50$

d) $n = 1\ 000, p = 7\ 500, k = 500$

Isomorphisme de graphes. Un isomorphisme de graphes colorés est une bijection entre les sommets de deux graphes à n sommets G_1 et G_2 qui préserve les arêtes et les couleurs. Nous ne considérerons dans ce sujet que des cas où G_1 et G_2 sont colorés par la même fonction de coloriage c. Plus formellement, on dit que f est un isomorphisme de graphe entre G_1 et G_2 si f est une fonction bijective telle que
— $\{k, \ell\} \in G_1$ si et seulement si $\{f(k), f(\ell)\} \in G_2$, et
— $c(k) = c(f(k))$ pour tout k.
On notera $G_1 \sim G_2$ si G_1 et G_2 sont isomorphes.
Le problème de l'isomorphisme de graphe, déterminant si deux graphes sont isomorphes, est un problème difficile. Dans ce sujet, nous allons donc nous intéresser à d'autres façons de juger si deux graphes colorés sont similaires.

2 Premières similarités

Dans cette section, nous considérons que les graphes $G_{n,p}$ et $\widehat{G}_{n,p}^k$ sont toujours colorés par la fonction de coloriage c^n définie dans les préliminaires.

Comparaison de fonctions. Au cours du sujet, il vous sera demandé de tester si des ensembles de fonctions sont identiques. Ces fonctions pourront toujours être vues comme étant de même support fini, et considérées égales si elles sont égales en tout point.

Couleurs des voisins. Étant donné un sommet i d'un graphe G à n sommets coloré par c, on note par $V_i^{G,c}$ la fonction associant à chaque couleur le nombre de voisins de i de cette couleur. Plus formellement, $V_i^{G,c} : [\![0, n-1]\!] \to [\![0, n-1]\!]$ est telle que $V_i^{G,c}(x) = |\{j \in [\![0, n-1]\!] : c(j) = x, \{i, j\} \in G\}|$, où $|\cdot|$ désigne le cardinal d'un ensemble.
Nous définissons alors la forme normale de voisinage de G comme l'ensemble $\{V_i^{G,c}\}_{i \in [\![0,n-1]\!]}$. Deux graphes G_1 et G_2 ont un voisinage similaire, noté $G_1 \equiv_V G_2$, s'ils ont la même forme normale de voisinage.

Question 4 *Étant données les valeurs de paramètres suivantes pour les variables n, p et k, calculer le cardinal de l'ensemble $\{i \mid G_{i,p} \equiv_V \hat{G}_{i,p}^k\}_{i \in [\![n,n+99]\!]}$.*

a) $n = 10, p = 500, k = 2$

b) $n = 100, p = 5000, k = 2$

Question à développer pendant l'oral 1 *Est-ce que $G_1 \equiv_V G_2$ si et seulement si $G_1 \sim G_2$? Prouver les implications correctes et fournir un contre-exemple aux implications incorrectes.*

Accessibilité. Soit G un graphe à n sommets coloré par c^n. Un sommet j est dit accessible depuis le sommet i s'il existe un chemin entre les deux sommets dans G. On prendra pour convention qu'un sommet est toujours accessible depuis lui-même (par le chemin vide). Pour tout sommet i, on note par A_i^G la fonction associant à chaque couleur le nombre de sommets accessibles depuis i de cette couleur. Plus formellement, $A_i^G : [\![0, n-1]\!] \to [\![0, n]\!]$ est telle que $A_i^G(x) = y$ si et seulement si y sommets accessibles depuis i ont la couleur x, c'est-à-dire que $|\{j \in [\![0, n-1]\!] \mid c(j) = x,\ j \text{ est accessible depuis } i\}| = y$.
Nous définissons alors la forme normale d'accessibilité de G comme l'ensemble $\{A_i^G\}_{i \in [\![0,n-1]\!]}$. Deux graphes G_1 et G_2 ont une accessibilité similaire, noté $G_1 \equiv_A G_2$, s'ils ont la même forme normale d'accessibilité.

Question 5 *Étant données les valeurs de paramètres suivantes pour les variables n, p et k, calculer le cardinal de l'ensemble $\{i \mid G_{i,p} \equiv_A \hat{G}_{i,p}^k\}_{i \in [\![n,n+99]\!]}$.*

a) $n = 10, p = 500, k = 2$

b) $n = 100, p = 300, k = 500$

Question à développer pendant l'oral 2 *Présenter votre algorithme et estimer sa complexité en fonction de n, p et k.*

Question à développer pendant l'oral 3 *Est-ce que $G_1 \equiv_A G_2$ si et seulement si $G_1 \sim G_2$? Est-ce que $G_1 \equiv_A G_2$ si et seulement si $G_1 \equiv_V G_2$? Prouver les implications correctes et fournir un contre-exemple aux implications incorrectes.*

3 Similarité d'arbres couvrants

Dans cette section, nous considérons que nos graphes $G_{n,p}$ et $\widehat{G}_{n,p}^k$ sont toujours colorés par la fonction de coloriage c^n définie dans les préliminaires.
Nous nous intéressons à présent à la notion d'arbre couvrant. Un arbre couvrant d'un graphe G depuis un sommet i est un graphe dirigé, connexe et sans cycle qui comprend tous les sommets accessibles depuis i et qui est un sous-graphe de G. Cet arbre admet i pour racine. En particulier, l'arbre construit en profondeur depuis i, noté T_i^G, est l'arbre couvrant qui s'obtient en parcourant le graphe en profondeur depuis i, tout en respectant l'ordre de visite des sommets suivant : lors de la visite d'un sommet i, si $\{i, j\}, \{i, k\} \in E^2$, $j < k$ et j et k n'ont pas encore été visités, alors j est visité avant k.
Étant donné un arbre T enraciné en i, nous nous intéressons aux couleurs des niveaux de cet arbre : un sommet j appartient au niveau k s'il y a un chemin de longueur k reliant la racine i au sommet j dans T. En particulier, le seul sommet au niveau 0 est i lui-même. On note par $N(T)$ la fonction associant à chaque niveau k de T et à chaque couleur c le nombre de sommets i présents dans T au niveau k et de couleur c.

Nous définissons alors la forme normale couvrante de G comme l'ensemble $\{N(T_i^G)\}_{i\in[\![0,n-1]\!]}$. Deux graphes G_1 et G_2 ont une couverture similaire, noté $G_1 \equiv_T G_2$, s'ils ont la même forme normale couvrante. Remarque : la relation de similarité de couverture est incomparable avec celle d'isomorphisme.

Question 6 *Étant données les valeurs de paramètres suivantes pour les variables n, p et k, calculer le cardinal de l'ensemble $\{i \mid G_{i,p} \equiv_T \hat{G}^k_{i,p}\}_{i\in[\![n,n+99]\!]}$.*

a) $n = 10, p = 500, k = 2$ ***b)*** $n = 25, p = 5\ 000, k = 2$

Question à développer pendant l'oral 4 *Présenter votre algorithme et estimer sa complexité en fonction de n et p.*

4 Similarité de Weisfeiler-Leman

Dans cette section, nous qualifions une fonction de coloriage $c : [\![0, n-1]\!] \to [\![0, n-1]\!]$ d'ordonnée si elle satisfait les deux propriétés suivantes :

1. $c(0) = 0$,
2. pour tout $i \in [\![0, n-1]\!]$, si $c(i) \neq c(j)$ pour tout $j < i$, alors $c(i) = \max\{c(j)\}_{j<i} + 1$.

Soit G un graphe à n sommets, coloré par une fonction de coloriage c. Nous définissons une nouvelle fonction de coloriage $\mathsf{next}(c)$ pour le graphe G comme étant la fonction de coloriage ordonnée satisfaisant la propriété suivante : pour tout sommets k et ℓ, $\mathsf{next}(c)(k) = \mathsf{next}(c)(\ell)$ si et seulement si

1. $c(k) = c(\ell)$ et
2. $V_k^{G,c} = V_\ell^{G,c}$

En d'autres termes, deux sommets ont la même couleur à l'étape suivante s'ils avaient déjà la même couleur et avaient la même quantité de voisins de chaque couleur.
Étant donnée la couleur initiale c d'un graphe G, on définit alors la suite de couleur $(c_k)_{k\in\mathbb{N}}$ telle que $c_0 = c$ et $c_{k+1} = \mathsf{next}(c_k)$ pour tout k.

Question à développer pendant l'oral 5 *Justifier que la suite $(c_k)_{k\in\mathbb{N}}$ est constante à partir d'un certain rang. On notera $\hat{c}_G(c_0)$ la couleur ainsi atteinte par $(c_k)_{k\in\mathbb{N}}$.*

On dit que deux sommets i et j appartiennent à la même classe de WL-équivalence s'ils ont la même couleur dans la forme normale de Weisfeiler-Leman, c'est-à-dire si $\hat{c}_G(c_0)(i) = \hat{c}_G(c_0)(j)$. Nous définissons alors la forme normale de Weisfeiler-Leman de G comme le multi-ensemble, c'est-à-dire l'ensemble avec répétitions, des cardinaux de ses classes de WL-équivalence. Formellement, c'est à dire le multi-ensemble des entiers $k > 0$, pour $x \in [\![0, n-1]\!]$, tel que $|\{i \mid \hat{c}_G(c_0)(i) = x\}| = k$. Deux graphes G_1 et G_2 initialement colorés par la même fonction de coloriage c_0 sont dit similaires de Weisfeiler-Leman, noté $G_1 \equiv^{c_0}_{WL} G_2$, s'ils ont la même forme normale de Weisfeiler-Leman.

Question 7 *Étant données les valeurs de paramètres suivantes pour les variables n, p et k, calculer le cardinal de l'ensemble $\{i \mid G_{i,p} \equiv^{c^n}_{WL} \hat{G}^k_{i,p}\}_{i\in[\![n,n+99]\!]}$.*

a) $n = 10, p = 500, k = 100$ ***b)*** $n = 30, p = 250, k = 100$

Question à développer pendant l'oral 6 *Présenter votre algorithme et estimer sa complexité en fonction de n et p.*

Question à développer pendant l'oral 7 *Considérons cst la fonction de coloriage uniforme associant 0 à chaque sommet. Est-ce que $G_1 \equiv^{cst}_{WL} G_2$ si et seulement si $G_1 \sim G_2$? Prouver les implications correctes et fournir un contre-exemple aux implications incorrectes.*

4.1 Coloriage et classes d'équivalences

Dans le reste de cette section, nous ne fixons pas le coloriage c_0 de nos graphes, elle est un paramètre que nous faisons varier.

Question 8 *Calculer le nombre maximal de classes de WL-équivalence que le graphe $G_{n,p}$ peut avoir vis-à-vis d'un coloriage initial c_0 d'au plus $n/2$ couleurs, pour les valeurs de paramètres suivantes.*

a) $n = 8, p = 5$

b) $n = 8, p = 20$

c) $n = 10, p = 5$

d) $n = 10, p = 50$

Question à développer pendant l'oral 8 *Présenter votre algorithme et estimer sa complexité en fonction de n et p.*

Fiche réponse type : Isomorphisme et similarités de graphes

$\widetilde{u_0}$: 42

Question 1

a) 182

b) 562

c) 489

d) 34

Question 2

a) (6, 6)

b) (2519, 69)

c) (20, 69)

d) (375230, 642)

Question 3

a) (1349, 1938)

b) (965980, 937887)

c) (348367, 842972)

d) (372841, 357944)

Question 4

a) 10

b) 12

Question 5

a) 73

b) 33

Question 6

a) 12

b) 30

Question 7

a) 78

b) 48

Question 8

a) 4

b) 4

c) 5

d) 8

Fiche réponse : Isomorphisme et similarités de graphes

Nom, prénom, u_0 : ....

Question 1

a)

b)

c)

d)

Question 2

a)

b)

c)

d)

Question 3

a)

b)

c)

d)

Question 4

a)

b)

Question 5

a)

b)

Question 6

a)

b)

Question 7

a)

b)

Question 8

a)

b)

c)

d)