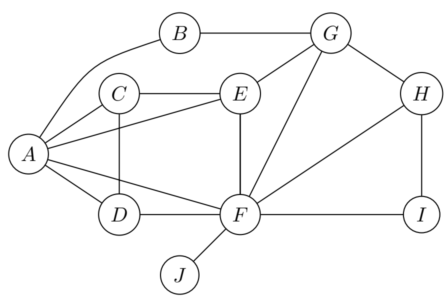
Projet : Dégénérescence des grands graphes réels

Associés au Big Data, les grands graphes (réseaux sociaux, infrastructure, biologie, etc) font l'objet d'une attention grandissante dans différents domaines. Si un graphe ne possède pas de sous-graphe dense, alors de nombreux calculs sur celui-ci deviennent plus faciles et rapides à opérer. De nombreuses façons de mesurer la densité d'un graphe ont été proposées, parmi elles la dégénérescence est la plus courante et on connaît un algorithme performant pour la calculer.

La notion de dégénérescence est liée à la notion de k-centre (k-*core* en anglais) : un k-*centre* d'un graphe non orienté G est un sous-graphe maximal H de G où tous les sommets sont de degré au moins k dans H. Le plus grand k pour lequel G contient un k-centre non nul est sa dégénérescence. Deux autres définitions équivalentes sont qu'un graphe a une dégénérescence de k si tout sous-graphe contient un sommet de degré inférieur ou égal à k ou bien s'il existe un ordre sur les sommets tel que tout sommet a au plus k voisins qui le précèdent dans cet ordre.

Il existe un algorithme simple pour calculer la dégénérescence d'un graphe : il suffit de partir de $k = 1$ et répéter la suppression des sommets de degré inférieur ou égal à k puis de passer à $k = 2$ et supprimer tous les sommets de degré au plus 2 et passer ensuite à $k = 3$ et ainsi de suite jusqu'à qu'il ne reste plus de sommet. Le dernier k pour lequel un sommet a été supprimé est la dégénérescence du graphe.

Exemple : Sur le graphe ci-dessous, pour $k = 1$, seul le sommet J sera supprimé ; pour $k = 2$ les sommets B, I, H seront supprimés, puis G (qui a maintenant un degré de 2) sera supprimés ; pour $k = 3$, les sommets C, D, E, F puis A seront supprimés, donc ce graphe a une dégénérescence de 3.



En plus de connaître la dégénérescence d'un graphe, il peut être intéressant de connaître, pour chaque sommet, son numéro de centre : il s'agit de la valeur de k au moment où il est supprimé. Sur le graphe ci-dessus, on a les numéro de centre suivants :

Sommet	A	B	C	D	E	F	G	H	I	J
Numéro de centre	3	2	3	3	3	3	2	2	2	1

On connaît des classes de graphes ayant une faible dégénérescence : les arbres ont une dégénérescence de 1, les graphes planaires ont une dégénérescence d'au plus 5.

L'objectif de ce projet est d'implanter l'algorithme de calcul de dégénérescence ci-dessus pour le tester sur des graphes réels (potentiellement assez grands) et ainsi vérifier en quelles mesures les graphes réels sont denses.

Travail à effectuer

En vous aidant des TP,

1. écrire une fonction pour coder le graphe G (le choix de la structure est à choisir pour une gestion optimale) à partir d'une description du graphe dans un fichier au format SNAP ou KONECT ;
2. écrire une fonction pour calculer la dégénérescence et stocker les numéro de centres dans un tableau pour un graphe donné (il faudra réfléchir à la meilleure façon de supprimer les sommets : est-ce une suppression réelle, ou juste un marquage du sommet pour indiquer qu'il est supprimé ou autre ?) ;
3. implanter au moins une des extensions suivantes :
   (a) algorithme de calcul optimisé de la dégénérescence de Matula & Beck ayant une complexité linéaire par utilisation du tri bucket (voir page Wikipédia).
   (b) comparer sur plusieurs graphes la dégénérescence et le nombre chromatique (en utilisant une estimation de celui-ci par DSATUR par exemple). En exécutant l'algorithme de coloration glouton avec un ordre des sommets suivant leur numéro de centre décroissant, on va colorier tout graphe k-dégénéré en au plus $k+1$ couleurs. Mais la différence entre $\chi(G)$ et la dégénérescence est-elle faible en moyenne ?
   (c) utiliser les numéros de centre pour obtenir un "joli" dessin du graphe : disposer les sommets de numéro 1 le long d'un cercle en périphérie ; puis les sommets de numéro 2 dans un cercle plus petit à l'intérieur, etc jusqu'aux sommets de plus grand numéro qui seront disposés en cercle à l'intérieur (le "noyau"). Une exportation en pdf sera bienvenue.

Graphes de tests

Pour tester vos algorithmes, un certain nombre d'instances de graphes sont disponibles en ligne : SNAP et KONECT

Modalités de réalisation

— Projet à réaliser en binôme, langage de programmation libre.
— La notation sera basée sur la façon dont les fonctions sont implantées, les résultats du programme ainsi que sur le mini-rapport.
— Le contenu attendu du rapport ainsi que les modalités d'évaluation du projet seront disponibles sur plubel.

Inscription

Questions techniques ou relatives à l'organisation

A poster sur le forum d'aide de plubel ou par mél olivier.togni@u-bourgogne.fr après avoir été voir si si la réponse n'apparait pas déjà sur le forum.