Grands Réseaux d'Interaction
Examen final
18 décembre 2018, 8h30-11h30

durée trois heures
le sujet est recto/verso
document autorisé : une feuille A4 manuscripte — tout autre document interdit

Exercice 1 : importance d'un sommet

Qu'est-ce qui permet de dire si un sommet est **important** dans un grand réseau d'interaction ? La question est large, on s'intéresse aussi bien à la définition formelle (comment poser en équation, par exemple, «central» ? «connu» ? «précieux» ? etc.) qu'au calcul (on ne demande pas de détailler des algorithmes mais de donnée l'idée en quelques lignes), et cela dépend de la nature du graphe. Inutile de demander, aucune précision supplémentaire ne sera donnée.

Exercice 2 : Erdős–Rényi contre réseau routier

Le but de cet exercice est de comparer une classe de graphes aléatoire avec un graphe réel :

- d'une part, le modèle $\mathcal{G}(n, \frac{4}{n})$, qui est une sous-classe de la classe des graphes aléatoires d'Erdős–Rényi $\mathcal{G}(n, p)$ obtenue en prenant $p = \frac{4}{n}$.
- d'autre part RRF, le réseau routier de la France, qui est un graphe avec un sommet par carrefour et une arête par route (non orienté : on ignore les sens interdits). Pour information il y a offciellement 1 073 500 kilomètres de routes en France, ce qui donne une idée ce ce que peuvent être n et m (il n'est pas demandé de deviner n et m)

Discutez des propriétés attendues (c'est-à-dire «en moyenne, en fonction de n le cas échéant» pour $\mathcal{G}(n, \frac{4}{n})$ et «qui semblent raisonnables» pour RRF) des graphes de ces deux classes pour les paramètres suivants

- en **donnant** quand c'est possible les deux valeurs,
- en les **comparant**, en supposant qu'on fixe n pour $\mathcal{G}(n, \frac{4}{n})$ à la même valeur que le nombre de sommets de RRF, il faut donc dire si c'est égal, du même ordre, ou qui est plus grand,
- et en **justifiant** vos réponses (avec un argument théorique pour $\mathcal{G}(n, \frac{4}{n})$ et empirique pour RRF)

1. degré moyen
2. distribution des degrés (précisez la valeurs des paramètres le plus possible)
3. diamètre
4. distance moyenne entre deux sommets (situés dans la même composante)
5. existence d'un coeur dense
6. existence d'une composante connexe géante
7. transitivité (coefficient de clustering)
8. navigabilité
9. existence de communautés

Exercice 3 : modéliser Facebook

Proposez un modèle de graphe aléatoire pour générer des réseaux sociaux, c'est-à-dire, des réseaux fabriqués avec de l'aléatoire et dont les propriétés caractéristiques (les neuf points de l'exercice précédent : degré moyen, etc.) soient les plus proches possibles de celles d'un réseau social réel. Détailler sous forme d'un algorithme : à chaque étape on ajoute ou enlève un ou plusieurs sommets ou arêtes, selon telles règles probabilistes. Cet algorithme peut se baser sur un modèle vu en cours (par exemple l'attachement préférentiel) mais doit y rajouter des choses. Justifier autant que possible (une preuve formelle qu'une propriété désirée est atteinte étant *a priori* très dure, on ne vous demande pas d'écrire une démonstration, mais de convaincre, en écrivant quelques lignes, que vous modèle est crédible)

Exercice 4 : comment trouver une donnée dans une DHT

On rappelle que dans une DHT chaque pair P a un numéro $n(P) \in [0, 2^{160}[$, chaque donnée D a un hash $h(D) \in [0, 2^{160}[$, et que chaque pair est responsable de l'indexation des données dont le hash est le plus proche de son numéro. Dire comment un pair peut trouver une donnée (de hash connu), d'une façon générale, puis en détaillant dans l'une des deux DHT vues en cours, Chord ou Kademlia (au choix l'une ou l'autre, ne pas traiter les deux). Donner le temps de la requête et le nombre de messages échangés.