

M2 – SDR (TP4-5)

Modèles de graphes

Clémence Magnien, Lionel Tabourier, Fabien Tarissan

Cette séance de TP vise à générer et décrire les propriétés des modèles de graphes vus en cours, pour les comparer à la structure des réseaux réels. Il sera donc nécessaire de savoir faire les mesures qui ont été réalisées au cours des premières séances de TP.

Dans la suite de la feuille (et les TPs suivants), nous représenterons naturellement les réseaux par des graphes, que nous noterons $G = (V, E)$, où V est l'ensemble des nœuds (aussi appelés *sommets*) du graphe numérotés de 0 à $|V| - 1$ et E l'ensemble des liens (aussi appelés *arêtes*).

1 Graphe aléatoire d'Erdős-Rényi

Exercice 1 — Génération. Réaliser un programme qui, étant donnés deux entiers positifs n et m , génère un graphe aléatoire d'Erdős-Rényi ayant n sommets et m arêtes et écrit la liste de ses arêtes dans un fichier `graphe.data` et son nombre de sommets dans `graphe.n`. On autorise les liens multiples, mais pas les boucles.

Exercice 2 — Caractéristiques. À l'aide des programme précédents :

1. Générer un graphe Erdős-Rényi avec $n = 7236$ sommets et $m = 22270$ arêtes.
2. Calculer les caractéristiques principales de ce graphe.
3. Afficher la distribution de degré de ce graphe à l'aide de GNUPLOT (cf. section 4.2)
4. Commenter ces résultats.

Exercice 3 — Comparaison au réseau PPI de la drosophile. Reprendre le réseau d'interaction protéine-protéine de la drosophile (que nous nommerons par la suite PPI-DROSO) vu au cours du premier TP (retirer les boucles si ce n'est pas déjà fait).

Reprendre les mesures réalisées au cours du premier TP sur PPI-DROSO, notamment sa distribution de degré et comparer ceux-ci avec ceux obtenu avec les réseaux Erdős-Rényi. Que peut-on déduire de la nature du réseau PPI-DROSO ? Quelle conclusions tirez-vous des valeurs du degré moyen pour chacun de ces deux réseaux ?

2 Graphe aléatoire à distribution de degrés fixés

Exercice 4 — Génération par la méthode directe. Réaliser un programme qui, étant donnée une liste de degrés, génère un graphe aléatoire respectant la distribution de degré de la liste.

Exercice 5 — Génération par la méthode d'échanges. Réaliser un programme qui, étant donné un graphe réel, effectue P permutations aléatoires de liens. On veillera à ce que les permutations de liens ne créent ni boucle, ni liens multiples.

Depuis le réseau PPI-DROSO de la drosophile, effectuer $P = 10^6$ permutations, et créer un fichier au format usuel contenant le graphe obtenu.

Ajouter au programme un module qui permet de mesurer le clustering du graphe toutes les 10^4 permutations. Écrivez dans un fichier la valeur du clustering à chacune des mesures, qu'observez-vous sur cette mesure au long du processus ?

Exercice 6 — Comparaison au réseau ppi-droso de la drosophile. Utiliser la méthode directe pour générer un graphe aléatoire ayant la même distribution de degré que PPI-DROSO.

1. Comparez les caractéristiques de ce nouveau graphe avec celles du graphe réalisé par la méthode d'échanges. Les réseaux sont-ils identiques ? Si non, qu'est-ce qui les différencie ?
2. Comparez les caractéristiques de ce graphe avec celles de PPI-DROSO. Les réseaux sont-ils identiques ? Si non, qu'est-ce qui les différencie ?

3 Générer un graphe de Barabási-Albert

Exercice 7 — Génération d'un graphe de Barabási-Albert. Réaliser un programme générant un graphe selon la méthode de Barabási-Albert, tel que :

1. le graphe soit construit par l'ajout de n sommets à un graphe de départ, ce graphe de départ sera entré en paramètre du programme
2. chaque sommet ajouté est lié à m nœuds du graphe existant selon la règle d'attachement préférentiel, m sera également un paramètre du programme

Exercice 8 — Comparaison au réseau ppi-droso de la drosophile. Utiliser l'algorithme précédent pour générer un graphe *scale-free* de Barabási-Albert, à $n = 7236$ sommets, à comparer au réseau PPI-DROSO, selon les indications suivantes :

1. prenez pour graphe de départ le graphe à 4 nœuds contenant les arêtes :
0 1
1 2
2 3
2. fixez la valeur de m pour avoir une densité du même ordre que celle du graphe PPI-DROSO.

Comparer la structure de PPI-DROSO à celle de ce graphe. Commentez vos observations.

4 Aide mémoire

4.1 Commande Unix

- `cut -f i,j fic` : extraction des champs i et j à partir d'un fichier fic
- `sort fic` : ordonne le fichier fic
- `uniq -c fic` : compte le nombre d'occurrence de lignes identiques consécutives (à combiner avec `sort` pour obtenir une distribution)
- `cmd > fic` : redirection de la sortie de la commande cmd dans le fichier fic .

4.2 Gnuplot

GNU PLOT est un programme permettant d'obtenir une représentation graphique à partir d'un jeu de données structuré. Beaucoup de documentations et tutoriels existent¹. On ne donne ici que la commande principale qui va vous être utile pour afficher les résultats. Il suffit d'exécuter la commande `gnuplot` dans un terminal, puis `plot "fic" using i:j`, ce qui conduira à afficher les données contenues dans le fichier fic en utilisant les valeurs de la colonne i pour abscisse et celles de la colonne j pour ordonnée.

Quelques options utiles :

- `set logscale x` : utilise une échelle logarithmique (et non linéaire) pour l'axe des x (*idem* avec y).
- `set xrange [v1 : v2]` : n'affiche les données que pour les valeurs de x comprises entre v_1 et v_2 (*idem* avec `yrange`).
- `set xlabel nom` : Donne le nom nom à l'axe des x (*idem* avec `ylabel`).

1. <http://www.gnuplot.info/>
<http://www.info.univ-angers.fr/~gh/tuteurs/tutgnuplot.htm>