Graphes - Rapport
Graphes Aléatoires

Renault Gabriel

Table des matières

1 Introduction

La compléxité des algorithmes se calcule souvent pour le pire cas. Mais pour les problèmes dits difficiles, il peut être intéressant de calculer la compléxité moyenne de certains algorithmes qui seraient lents sur quelques entrées et bien plus rapides sur une grande majorité d'entre elles. Pour cela, il faut utiliser une analyse probabiliste qui regardera la probabilité d'occurence des différentes entrées, et qui donnera le temps moyen d'éxécution de l'algorithme en tenant compte de cette distribution.

Ici, nous allons nous intéresser à deux modèles de distribution.

2 Premier Modèle : Modèle d'Erdös Renyi

2.1 Présentation

On choisit p une fonction de $\mathbb{N}$ dans $[0;1]$. On note n le nombre de sommets du graphe que l'on va générer et on considère que chaque arête de ce graphe a une probabilité $p(n)$ d'être présente.

Étant donnée une telle fonction p, on dira qu'une propriété est vraie pour presque tous les graphes si la probabilité qu'elle soit vraie sur un graphe généré par cette méthode tend vers 1 lorsque le nombre de sommets tend vers l'infini. On pourra aussi dire que cette propriété est presque toujours vraie.

2.2 Propriétés

2.2.1 Propriétés démontrées

Lorsque p est une fonction constante dont la valeur d'arrivée, qu'on notera abusivement p, n'est ni 0 ni 1, presque tous les graphes ont un diamètre 2.

Montrons d'abord que le diamètre est presque toujours inférieur ou égal à 2. Soient x et y deux sommets distincts du graphe.
Pour tout sommet z du graphe distinct de x et de y, il y a une probabilité $(1 - p^2)$ qu'au moins une des deux arêtes (x, z) et (y, z) ne soit pas présente dans le graphe.
Il y a $(n-2)$ sommets de ce type, donc une probabilité $(1-p^2)^{n-2}$ que pour tout sommet z du graphe distinct de x et de y, une des deux arêtes (x, z) et (y, z) ne soit pas présente dans le graphe, puisque ces évènements sont indépendants.
Il faut considérer moins de n^2 paires de sommets, donc la probabilité qu'il existe une paire de sommets dont la distance soit strictement supérieure à 2 est inférieure à $n^2(1 - p^2)^{n-2}$ (ici, on a majoré grossièrement en faisant la somme des probabilités car on ne peut pas faire vraiment mieux sans regarder de plus près puisqu'elles ne sont pas indépendantes, et en ne prenant pas en compte le cas où la paire de sommets est reliée par une arête) qui tend bien vers 0 lorsque n tend vers ∞.

Montrons maintenant que le diamètre n'est presque jamais 1.
S'il l'était, c'est que toutes les arêtes ont été générées, ce qui arrive avec probabilité $p^{\frac{n(n-1)}{2}}$ qui tend bien vers 0 lorsque n tend vers ∞.

On a bien la propriété recherchée.

Soit H un graphe quelconque fixé. Lorsque p est une fonction constante dont la valeur d'arrivée, qu'on notera abusivement p, n'est ni 0 ni 1, le graphe contient presque toujours H comme sous-graphe induit.

On note k le nombre de sommets de H et l le nombre de ses arêtes.
Pour tout ensemble de k sommets dans G, la probabilité d'être isomorphe à H est supérieure à $p' = p^l(1-p)^{\frac{k(k-1)}{2}-l}$ (on numérote les sommets au hasard et on regarde si cela correspond, bien entendu, il est possible qu'il soit isomorphe d'une autre façon mais on ne s'en occupe pas).
On peut créer $\lfloor \frac{n}{k} \rfloor$ tels ensembles disjoints entre eux, ce qui donne une probabilité supérieure à $1-(1-p')^{\lfloor \frac{n}{k} \rfloor}$ au graphe de contenir H comme sous-graphe induit, qui tend bien vers 1 lorsque n tend vers ∞.

On a bien la propriété recherchée.

Lorsqu'il existe un réel strictement positif α tel que $p(n) = \min(1, \frac{\alpha}{n})$, le graphe n'est presque jamais connexe.

Soient x et y deux sommets distincts du graphe.
On va montrer que ces deux sommets ne sont presque jamais reliés.
On se place directement dans le cas où n est supérieur à α.
Pour tout entier k compris entre 1 et $n-1$, il existe C_{n-2}^{k-1} chemins passant par k arêtes ne passant pas plusieurs fois par le même sommet, et chacun d'eux a une probabilité d'être intégralement présent égale à $(\frac{\alpha}{n})^k$.
On peut donc majorer grossièrement la probabilité que x et y soient reliés par la somme de ces probabilités.
Or :

$$\begin{aligned}
\sum_{k=1}^{n-1} C_{n-2}^{k-1} * \frac{\alpha}{n}^k &\leqslant \sum_{k=1}^{\infty} C_{n-2}^{k-1} * \frac{\alpha}{n}^k \\
&\leqslant \sum_{k=1}^{\infty} C_{n}^{k-1} \frac{\alpha}{n}^k \\
&\leqslant \sum_{k=1}^{\infty} \frac{n^{k-1}}{(k-1)!} \frac{\alpha}{n}^k \\
&\leqslant \sum_{k=1}^{\infty} \frac{\alpha^k}{(k-1)!} \frac{1}{n} \\
&\leqslant \frac{1}{n} \sum_{k=1}^{\infty} \frac{\alpha^k}{(k-1)!}
\end{aligned}$$

Or $\frac{\alpha^k}{(k-1)!} \leqslant \frac{1}{k^2}$ à partir d'un certain rang donc la série converge vers une constante que nous nommerons c_α.
La probabilité que x et y soient reliés est donc inférieure à $\frac{c_\alpha}{n}$ qui tend bien vers 0 lorsque n tend vers ∞.

On a bien la propriété recherchée.

2.2.2 Propriétés admises

Lorsque p est une fonction constante dont la valeur d'arrivée, qu'on notera abusivement p, n'est ni 0 ni 1, et qu'on fixe $\epsilon > 0$, presque tous les graphes ont une clique max de taille entre $\lfloor 2\log_{\frac{1}{p}} n - \log_{\frac{1}{p}} \log_{\frac{1}{p}} n + 1 + 2\log_{\frac{1}{p}} \frac{e}{2} - \epsilon \rfloor$ et $\lfloor 2\log_{\frac{1}{p}} n - \log_{\frac{1}{p}} \log_{\frac{1}{p}} n + 1 + 2\log_{\frac{1}{p}} \frac{e}{2} + \epsilon \rfloor$.

Lorsque p est une fonction constante dont la valeur d'arrivée, qu'on notera abusivement p, n'est ni 0 ni 1, et qu'on fixe $\epsilon > 0$, le nombre chromatique de presque tous les graphes est entre $\frac{(\frac{1}{2}-\epsilon)n}{\log_{\frac{1}{1-p}} n}$ et $\frac{(1+\epsilon)n}{\log_{\frac{1}{1-p}} n}$.

3 Second Modèle

3.1 Présentation

Dans ce modèle, on va générer aléatoirement la séquence des degrés des sommets avec des fonctions génératrices. On definit

$$G_0(x) = \sum_{k=0}^{\infty} p_k x^k$$

où p_k est la probabilité qu'un sommet fixé ait pour degré k. On assure que la somme des p_k est bien égale à 1. On choisit ensuite un graphe uniformément parmi ceux qui ont la séquence de degrés selectionnée.
Dans la définition, on va jusqu'à l'infini parce qu'a priori, ce n'est pas borné mais en général, les p_k sont nuls à partir d'un certain rang.

Dans certains cas, les graphes étudiés sont toujours bipartis. On détermine alors deux fonctions génératrices f_0 et g_0 qui détermineront les propriétés pour chacune des parties du graphe.

Pour les graphes orientés, on considère la fonction génératrice

$$\mathcal{G}(x, y) = \sum_{k=0}^{\infty} p_{kj} x^k y^j$$

où p_{kj} représente la probabilité pour un sommet d'avoir un degré entrant k et un degré sortant j.

L'espérance du degré d'un sommet est

$$\langle k \rangle = \sum_{k=0}^{\infty} k p_k = G_0'(1).$$

3.2 Exemples de distributions

Soit p un reél strictement compris entre 0 et 1. On définit

$$p_k = C_n^k p^k (1-p)^{n-k}$$

où n est le nombre de sommets du grapheénéré.

$$\begin{aligned} G_0(x) &= \sum_{k=0}^{\infty} C_n^k p^k (1-p)^{n-k} x^k \\ &= \sum_{k=0}^{n} C_n^k p^k (1-p)^{n-k} x^k \\ &= (1-p+px)^n. \\ G_0'(x) &= np(1-p+px)^{n-1}. \end{aligned}$$

Ici, l'espérance du degré d'un sommet est

$$\langle k \rangle = G_0'(1) = np(1-p+p1)^{n-1} = np.$$

Pour certains graphes dont on connaît à l'avance la distribution, c'est à dire le nombre n_k de sommets de degré k, on définit

$$p_k = \frac{n_k}{\sum_{l=0}^{\infty} n_l}.$$

$$G_0(x) = \frac{\sum_{k=0}^{\infty} n_k x^k}{\sum_{k=0}^{\infty} n_k}.$$

4 Commentaires

Les propriétés admises sont démontrées dans le livre de West *Introduction to Graph Theory*, et c'est pourquoi j'ai préféré mettre ici les démonstrations laissées en exercice dans ce même recueil.

5 Bibliographie

Introduction to Graph Theory, Douglas B. West

Random graphs with arbitrary degree distributions and their applications, M.Newman, S.Strogatz et D.Watts