

Graphes aléatoires

Thomas Braibant

Introduction

Si l'on se place dans la perspective de l'analyse d'un algorithme, la complexité dans le pire cas est une mesure pessimiste de la qualité d'un algorithme. Tout comme certains problèmes sont dans NP à cause d'une petite famille d'instances dures, certains algorithmes peuvent être performants sur la plupart des instances, et être moins efficaces sur des instances *bizarres*.

L'analyse probabiliste, qui consiste à fixer une distribution de probabilité sur les entrées, et étudier l'espérance (ou valeur moyenne), du coût d'un algorithme en fonction de cette distribution donne une mesure plus acceptable. Néanmoins, choisir une distribution de probabilité réaliste peut s'avérer difficile. Dans ce rapport, on va donc étudier deux formalisations différentes de distributions de probabilités sur les graphes ayant un nombre de sommets fixés. Une fois le formalisme établi, on analysera ensuite :

- Quelques propriétés et leurs conditions d'apparition
- Un exemple d'algorithme polynomial qui permet de répondre pour presque tous les couples de graphes à la question de savoir si ils sont isomorphes
- Des comparaisons entre les prédictions et les données expérimentales tirées des graphes *réels* : Internet, réseaux divers, réseaux de collaboration (Acteurs -films, scientifique ...)

Les sources dont je me suis inspiré pour la rédaction de ce rapport sont les suivantes :

- [?] :)
- [?] : Un livre spécialement consacré aux graphes aléatoires, contenant beaucoup de résultats complexes sur le modèle A.
- [?] : Un revue bibliographique sur le sujet des graphes aléatoires, cité environ 1200 fois¹
- [?] : Un article présentant le modèle B. Pour information, les auteurs ne sont pas informaticiens, mais plutôt issus des milieux des sciences sociales ou de la matière. Article cité 250 fois environ²

1 Modélisations

J'ai regroupé quelques informations utiles en annexe, sur les probabilités, et quelques théorèmes. Le lecteur peu familier (comme moi) avec les notions de probabilités peut s'y référer.

1.1 Modèle A

Modèle A : Erdős-Renyi Soit un ensemble $V = \{1, \dots, n\}$ de sommets. Soit un paramètre $p, 0 < p < 1$. Pour toute paire de sommets, il existe une arête avec probabilité p , les arêtes étant déterminées indépendamment les unes des autres. La variable aléatoire G^p dénote un graphe issu de cet espace de probabilités.

¹information basée sur le site www.citebase.org

²idem

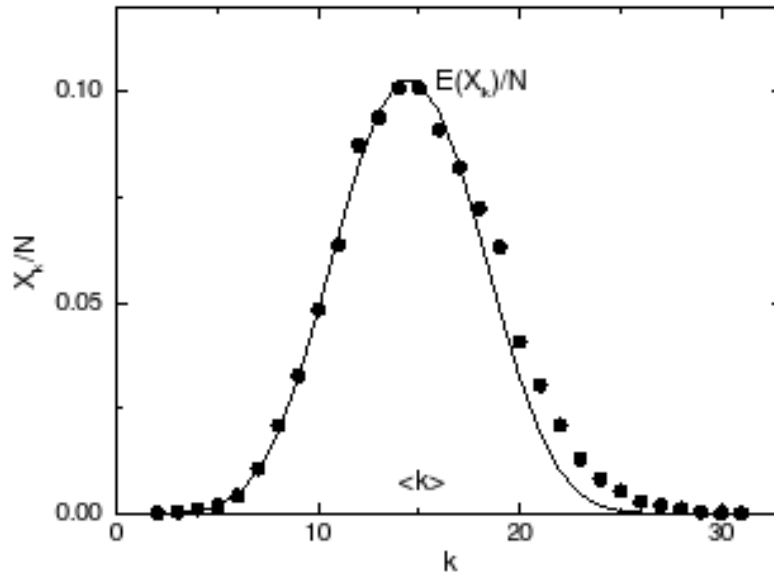


FIG. 1 – Distribution des degrés , figure extraite de [?], page 12

Lemme 1.1 (Degré d'un sommet) *Le degré d'un sommet est défini par la loi de probabilité (où on reconnaît une loi de Poisson) :*

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

Le degré moyen d'un sommet est donc :

$$z = E(k) = p * (n - 1)$$

La figure ?? indique les prévisions du modèle théorique correspondent à ce que l'on observe si l'on génère un graphe suivant cette modélisation. J'ai personnellement réalisé des tests sur un graphe aléatoire avec 5000 noeuds, et trouvé une distribution des degrés correspondant également à ces prévisions, aux erreurs expérimentales près.

Autre formalisation Soit n et $m = m(n)$. On définit la variable aléatoire G^m comme étant un graphe à n sommets et m arrêtes, apparaissant avec une probabilité $\binom{n}{m}^{-1}$. La variable aléatoire G^m dénote un graphe issu de cet espace de probabilités.

Cette seconde définition est plus commode dans de nombreux modèles étudiés, étant donné qu'il est plus facile de se demander, à n fixé, combien d'arêtes sont nécessaires pour avoir presque toujours une propriété Q , plutôt que de se demander quelle est la probabilité p nécessaire pour avoir presque toujours Q .

Ce second modèle est fort heureusement décrit par le modèle A, quand n est grand, et $p = m/\binom{n}{2}$.

1.2 Modèle B

De nombreuses sources convergentes indiquent que la distribution des degrés dans les graphes réels étudiés (systèmes biologiques, réseaux sociaux, réseaux métaboliques, ...) ne suit pas la

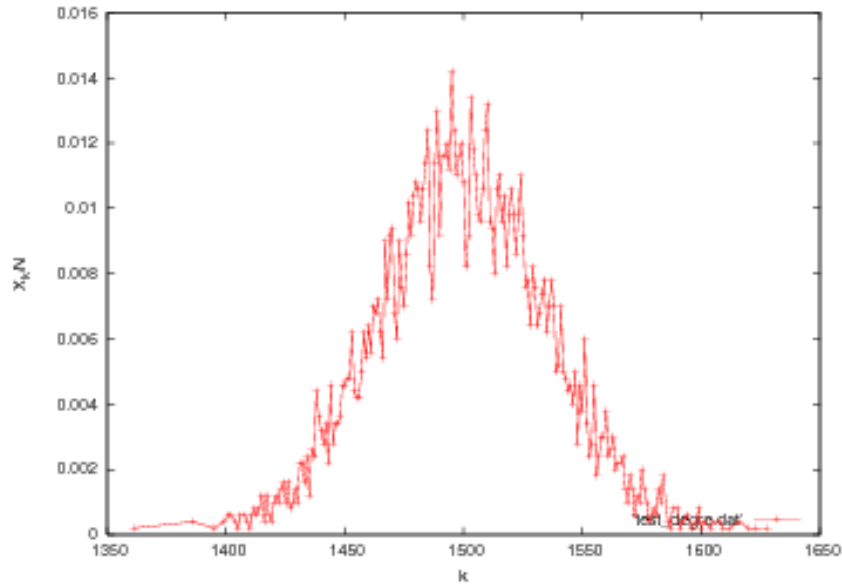


FIG. 2 – Distribution des degrés, expérience avec 5000 noeuds

loi de Poisson ci-dessus, et que de nombreuses propriétés sont mal approximées si on utilise le modèle A. On verra dans la section ?? des données expérimentales, mais il convient de dire en particulier que la répartition des degrés suit souvent une loi de puissance, qui n'est pas simulable dans le modèle d'Erdős-Renyi.

Définition On définit la fonction génératrice $G_0(x)$ telle que

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

avec p_k la probabilité qu'un sommet choisi au hasard soit de degré k . On ne travaillera qu'avec G_0 normalisée telle que $G_0(1) = 1$. La variable aléatoire $G^{(p_k)}$ dénote un graphe issu de cet espace de probabilités.

Cette modélisation permet de calculer de façon simple certaines quantités.

Lemme 1.2 (Degré moyen d'un sommet) *Le degré moyen d'un sommet (ou espérance) est donné par :*

$$z = \sum_k k * p_k = G_0'(1)$$

Lemme 1.3 (Puissances) *Si la loi de distribution d'une propriété X sur un objet est donnée par une certaine fonction génératrice, alors la loi de distribution sur la somme de X sur m éléments indépendants est donnée par la m -ième puissance de cette fonction génératrice.*

Somme des degrés de deux sommets pris au hasard

$$\begin{aligned} [G_0(x)]^2 &= \left[\sum_k p_k * x^k \right]^2 \\ &= \sum_{j,k} p_k * p_j * x^{j+k} \end{aligned}$$

Les coefficients de x^n dans cette expression sont clairement les produits $p_j * p_k$ tels que $j + k = n$, et par conséquent, on exprime bien ici la probabilité que la somme des degrés des deux sommets soit n .

1.3 Relations entre ces modèles

Si il n'existe pas d'arêtes multiples entre un même couple de sommets, il est clair que le modèle G^p recouvre le modèle G^m . Dans le cas contraire, il n'est pas facile de voir comment simuler cette possibilité.

Le modèle B recouvre le modèle A, si on utilise comme probabilité pour le degré k : $p_k = \binom{n}{k} p^k (1-p)^{n-k}$, la probabilité donnée en ??.

2 Propriétés

2.1 Propriétés pour presque tous les graphes

Définition Etant donnée une suite d'espaces de probabilités Ω_n , soit q_n la probabilité que la propriété Q soit vraie dans l'espace Ω_n . On dit que la propriété Q est presque toujours vérifiée, si $\lim_{n \rightarrow \infty} q_n = 1$

2.2 Diamètre

On rappelle que le diamètre d'un graphe est défini par : voir en annexe.

Théorème 2.1 (Diamètre) *Si p est constant, alors presque tous les graphes G^p (on se place ici dans le cas du Modèle A) ont pour diamètre 2.*

Preuve Soit $X(G^p)$ le nombre de paires de sommets sans voisins communs. Si il n'y en a pas, alors G^p est connecté, et a pour diamètre 2. On pose $X = \sum_{i,j \in [1..n]^2} X_{i,j}$ avec $X_{i,j} = 1$ si et seulement si v_i, v_j n'ont pas de voisins communs. Soient i_0, j_0 tels que $X_{i_0, j_0} = 1$. Les sommets $n - 2$ autres sommets n'ont pas d'arêtes vers i_0 et j_0 à la fois. On en déduit $P[X_{i,j} = 1] = (1 - p^2)^{n-2}$. Donc $E(X) = \binom{n}{2} * (1 - p^2)^{n-2}$.

Si p est fixé, alors $E(X) \mapsto 0$. Par l'inégalité de Markov, rappelée en annexe, cela implique que $P(X = 0) \mapsto 1$ et donc presque tous les graphes G^p ont pour diamètre 2.

Fin de preuve

Contrairement à ce qu'on peut penser, ce résultat est en réalité assez faible. En effet, dans les graphes usuels, p est une fonction décroissante du nombre de sommets³. On trouve donc souvent dans la littérature des résultats qui mettent en relation l'apparition d'une propriété avec un comportement asymptotique de p , de la forme suivante.

Théorème 2.2 (Diamètre) *Dans le cadre du modèle A, on a :*

- Si $p^2 * n - 2 * \log n \mapsto \infty$ et $n^2(1 - p) \mapsto \infty$ alors presque tous les graphes G^p ont pour diamètre 2.
- Si la fonction $M = M(n) < \binom{n}{2}$ vérifie $2 * \frac{M^2}{n^3} - \log n \mapsto \infty$ alors presque tous les graphes G^M ont pour diamètre 2

Preuve [?], page 263

Fin de preuve

³Soit X le nombre d'arêtes : $E(X) = \binom{n}{2} * p(n)$

2.3 Fonctions de seuils

Fonction de seuil On appelle *fonction de seuil de probabilité* pour une propriété monotone Q^4 une fonction $t(n)$ telle que $p(n)/t(n) \mapsto 0$ implique que presque aucun G^p ne vérifie la propriété Q , et telle que $p(n)/t(n) \mapsto \infty$ implique que presque tous les G^p vérifient Q .

Cette formalisation de la notion précédente nous amène à réaliser des analyses plus fines des propriétés des graphes aléatoires. Par exemple, dans le modèle A, $t : n \mapsto \ln(n)/n$ est une fonction de seuil de probabilité pour la disparition des sommets isolés.

3 Variables aléatoires

Dans cette section, je vais essayer de présenter des résultats permettant de mettre en relation des variables aléatoires identiques dans les deux modèles de graphes.

3.1 Distance moyenne

Il est clair que dans les deux modèles, le nombre moyen de premiers voisins est égal au degré moyen dans le graphe. Dans le modèle B, on a :

$$z_1 = \langle k \rangle = \sum_k k * p_k = G'_0(1)$$

($\langle k \rangle$ désignant l'espérance du degré)

Si on part d'un noeud aléatoire n , on s'intéresse alors à la distribution des degrés pour les premiers voisins atteints. Une arête quelconque prise au hasard atteint un noeud n' de degré k avec une probabilité proportionnelle à $k * p_k$. On a donc :

$$G_1(x) = \frac{\sum_k k * p_k * x^{k-1}}{\sum_k k * p_k} = \frac{1}{\langle k \rangle} G'_0(x)$$

Le nombre moyen de voisins est $z_1 = \langle k \rangle = G'_0$.

Par extension, on appelle z_m le nombre de m -ièmes voisins, et on a :

$$z_m = [G'(1)]^{m-1} G'_0(1) = \left[\frac{z_2}{z_1} \right]^{m-1}$$

Si on suppose que tous les noeuds sont accessibles en D étapes, on a :

$$1 + \sum_{m=1}^t z_m = n$$

D'où on a :

$$l = \frac{\ln(n/z_1)}{\ln(z_2/z_1)} + 1$$

Dans le modèle A, on peut raisonnablement penser que le nombre de sommets atteints à distance l en partant d'un noeud quelconque est de l'ordre de $\langle k \rangle^l$. En posant $\langle k \rangle^l = n$, il vient :

$$l \propto \frac{\ln(n)}{\ln(\langle k \rangle)}$$

⁴voir en annexe ??

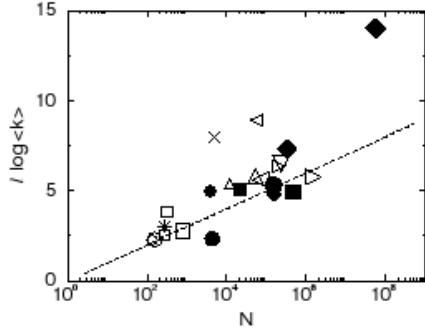


FIG. 3 – Longueur moyenne de chemin, comparaison entre modélisation et graphes réels

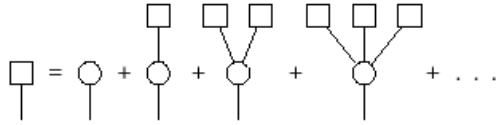


FIG. 4 – Une composante connexe peut être vue comme un sommet relié à différentes composantes connexes

3.2 Taille moyenne de composante connexe

Je vais ici décrire un résultat général pour le modèle B, et le raffiner ensuite pour le modèle A.

Dans le modèle B, on s'intéresse à la fonction génératrice $H_1(x)$ pour la taille des composantes atteintes en suivant une arrête jusqu'à son bout, en excluant le cas où le graphe possède une composante géante. La probabilité d'avoir un cycle étant en $1/N$, on peut la négliger pour N grand.

Remarquons tout d'abord avec la loi de puissance⁵ que la fonction génératrice de la taille de la composante formée de k sous composantes distinctes s'exprime comme la puissance k -ième de $H_1(x)$.

Donc, en s'appuyant sur le schéma figure ??, on remarque que $H_1(x)$ doit vérifier l'équation suivante, avec q_i la probabilité que le sommet initial aie i arrêtes sortantes :

$$H_1(x) = x * q_0 + x * q_1 * H_1(x) + x * q_2 * [H_1(x)]^2 + \dots$$

D'où il vient :

$$H_1(x) = xG_1(H_1(x))$$

En partant d'un sommet plutôt que d'une arrête, il vient

$$H_0(x) = xG_0(H_1(x))$$

De façon usuelle, on a alors :

$$\begin{aligned} \langle s \rangle &= H_0'(1) = 1 + G_0'(1) * H_1'(1) \\ &= 1 + \frac{G_0'(1)}{1 - G_1'(1)} \\ &= 1 + \frac{z_1^2}{z_1 - z_2} \end{aligned}$$

⁵(voir ??)

Notons d'ailleurs que cette équation peut diverger, et donne une condition pour l'apparition d'une composante géante, qui est :

$$\sum_k k * (k - 2) * p_k = 0$$

Pour le modèle A, on a $G_0(x) = G_1(x) = \exp(n - 1) * p * (x - 1)$, d'où $\langle s \rangle = 1 + \frac{(n-1)*p}{1-(n-1)*p}$.

3.3 Coefficient de clusterisation

Lemme 3.1 *Dans un graphe aléatoire G^p , $C_p = p = \frac{\langle k \rangle}{n-1}$.*

En effet, dans le modèle A, la probabilité qu'il existe une arête entre deux voisins d'un même noeud est la même que la probabilité d'existence d'une arête dans le cas général.

Je reprends ici une explication que m'a donnée Eric Thierry sur le modèle B.

Soit $A_{x,y}$ la variable aléatoire entière valant 1 si il existe une arête entre x et y et 0 sinon. Le modèle étant symétrique par rapport aux sommets, $P[A_{x,y} = 1]$ est identique pour tout couple $x, y \in V \times V, x \neq y$. On la note p .

$$\begin{aligned} p &= \frac{\sum_{x,y} P[A_{x,y} = 1]}{\text{nombre de paires } x,y} \\ &= \frac{E(\sum_{x,y} A_{x,y})}{n * (n - 1) / 2} \\ &= \frac{E(\text{nombre d'arêtes})}{n * (n - 1) / 2} \\ &= \frac{E(\text{somme des degrés})}{n * (n - 1)} \\ &= \frac{n * \langle k \rangle}{n * (n - 1)} \\ p &= \frac{\langle k \rangle}{n - 1} \end{aligned}$$

Comme toujours en étudiant des probabilités, la solution est évidente, lorsqu'on l'a trouvée. Ce résultat implique cependant que le coefficient de clusterisation est donné par la même expression dans les deux modèles, ce qui est intuitif. Attention, cependant, car il est possible de construire presque arbitrairement une distribution de degré donnant une valeur donnée de ce coefficient p pour le modèle B, tout en n'ayant que deux types de sommets de degrés a et b , et en choisissant judicieusement leurs probabilités respectives.

4 Applications

Les propriétés des graphes aléatoires peuvent permettre de trouver des algorithmes qui résolvent des problèmes difficiles sur presque toutes les entrées. Le problème de décision de l'isomorphisme de deux graphes est difficile⁶. Cependant, il est possible de construire un algorithme rapide qui permet de résoudre ce problème sur presque tous les graphes.

L'idée de construction de cet algorithme repose sur certaines propriétés de la séquence des degrés d'un graphe aléatoire.

⁶On ne sait pas si il est NP-Complet ou non

Théorème 4.1 (Erdős-Rényi) Soient $p = w_n \log(n)/n$ et $\epsilon > 0$ fixés, avec $n \mapsto w_n$ une fonction non bornée à croissance arbitrairement lente. Alors presque tous les graphes G^p vérifient :

$$(1 - \epsilon) * p * n < \delta(G^p) \leq \Delta(G^p) \leq (1 + \epsilon) * p * n$$

Théorème 4.2 (Bollobás) Dans le modèle A, avec p fixé, et $t(n) = \theta((n/\log(n))^{1/4})$, les $t(n)$ sommets de plus haut degré sont de degrés différents pour presque tout G^p .

Ce résultat va nous permettre de construire un algorithme permettant de tester l'isomorphisme de deux graphes de façon *rapide*.

Théorème 4.3 (Babai-Erdős-Selkow) Il existe un algorithme quadratique qui teste l'isomorphisme de presque toute paire de graphes.

Preuve L'idée sous-jacente est la suivante : on construit un ensemble \mathbb{H} qui contient presque tous les graphes, et on montre que l'isomorphisme avec un graphe de \mathbb{H} peut être testée de façon rapide.

Construction de \mathbb{H} Soit un graphe G à n sommets, codé par matrices d'adjacence. Calculer le degré de ses sommets, et les indiquer par degré décroissant. On pose $r = \lfloor 3 * \log(n) \rfloor$. Si $d(v_i) = d(v_{i+1})$ pour un $i < r$, on rejette G . Le théorème ci dessus montre avec $p = 1/2$ que presque tous les graphes passent ce test.

On pose $U = v_1, \dots, v_r$. Il y a de l'ordre de n^3 différents sous ensembles de U . On construit \mathbb{H} comme l'ensemble des graphes non rejetés, tels que les sommets de $V - U$ ont des voisinages différents dans U . Cela peut être fait en $\Theta(n^2)$ étapes : pour tout $x \in V - U$, on encode $N(x) \cap U$ comme un r -uplet binaire. On peut donc les trier comme des nombres binaires en $\Theta(n \log n)$. On renomme les sommets restants dans l'ordre décroissant de ces valeurs, en éliminant le graphe si deux valeurs consécutives sont identiques.

Il faut encore montrer que pour presque tous les graphes G^p la probabilité pour les sommets de $V - U$ d'avoir le même voisinage dans U tend vers 0. On peut montrer que le nombre de paires de sommets en dehors de U avec des voisinages identiques dans U est bornée par $O(\binom{n-r}{2}(1-p)^r)$. Or cette quantité tend vers 0.

Utilisation de \mathbb{H} Les graphes de \mathbb{H} sont les graphes qu'il est facile d'étiquetter. Supposons maintenant qu'on aie deux graphes étiquetés dans \mathbb{H} , isomorphes au sens de l'étiquettage. Il suffit alors de comparer les matrices d'adjacences, dont les lignes et les colonnes sont indicées par les étiquettes. Deux graphes sont isomorphes si et seulement si les matrices sont identiques, ce qui se vérifie en $\Theta(n^2)$ étapes.

Fin de preuve

5 Expériences

La figure ?? indique les caractéristiques générales de nombreux graphes. Pour chaque graphe, certaines quantités théoriques (longueur moyenne de chemin, coefficient de clusterisation) sont comparées aux prévisions théoriques.

On remarque que la longueur moyenne de chemin est assez bien décrite par le modèle théorique A. Par contre le coefficient de clusterisation est beaucoup plus important que les prévisions théoriques du modèle A. [?], page 14 indique que même pour le modèle B, il existe parfois un facteur 2 entre prévision et réalisation.

Network	Size	$\langle k \rangle$	ℓ	ℓ_{rand}	C	C_{rand}	Reference	Nr.
WWW, site level, undir.	153,127	35.21	3.1	3.35	0.1078	0.00023	Adamic 1999	1
Internet, domain level	3015 - 6209	3.52 - 4.11	3.7 - 3.76	6.36 - 6.18	0.18 - 0.3	0.001	Yook <i>et al.</i> 2001a, Pastor-Satorras <i>et al.</i> 2001	2
Movie actors	225,226	61	3.65	2.99	0.79	0.00027	Watts, Strogatz 1998	3
LANL coauthorship	52,909	9.7	5.9	4.79	0.43	1.8×10^{-4}	Newman 2001a,b	4
MEDLINE coauthorship	1,520,251	18.1	4.6	4.91	0.066	1.1×10^{-5}	Newman 2001a,b	5
SPIRES coauthorship	56,627	173	4.0	2.12	0.726	0.003	Newman 2001a,b,c	6
NCSTRL coauthorship	11,994	3.59	9.7	7.34	0.496	3×10^{-4}	Newman 2001a,b	7
Math coauthorship	70,975	3.9	9.5	8.2	0.59	5.4×10^{-5}	Barabási <i>et al.</i> 2001	8
Neurosci. coauthorship	209,293	11.5	6	5.01	0.76	5.5×10^{-5}	Barabási <i>et al.</i> 2001	9
<i>E. coli</i> , substrate graph	282	7.35	2.9	3.04	0.32	0.026	Wagner, Fell 2000	10
<i>E. coli</i> , reaction graph	315	28.3	2.62	1.98	0.59	0.09	Wagner, Fell 2000	11
Ythan estuary food web	134	8.7	2.43	2.26	0.22	0.06	Montoya, Solé 2000	12
Silwood park food web	154	4.75	3.40	3.23	0.15	0.03	Montoya, Solé 2000	13
Words, cooccurrence	460,902	70.13	2.67	3.03	0.437	0.0001	Cancho, Solé 2001	14
Words, synonyms	22,311	13.48	4.5	3.84	0.7	0.0006	Yook <i>et al.</i> 2001	15
Power grid	4,941	2.67	18.7	12.4	0.08	0.005	Watts, Strogatz 1998	16
<i>C. Elegans</i>	282	14	2.65	2.25	0.28	0.05	Watts, Strogatz 1998	17

FIG. 5 – Figure extraite de [?], page 8, comparant les données issues de graphes réels aux prévisions statistiques

Conclusion

En conclusion de ce rapport, je dois dire que le sujet des graphes aléatoires est extrêmement riche et intéressant. Les différentes modélisations peuvent être mises en examen en les comparant aux données expérimentales, ce qui conduit actuellement à un regain d'intérêt pour ce domaine.

Les modélisations présentées ici sont des modélisations parmi d'autres, en particulier, elles ne traitent pas des problèmes de graphes orientés, et pourraient être raffinées pour fournir des résultats plus intéressants sur des problèmes de graphes bipartis, petit-monde ou autres.

Un aspect que je n'ai pas du tout abordé, mais qui est évoqué de façon importante dans les sources que j'ai parcourues est l'évolution temporelle des graphes auxquels on ajoute ou retranche des arêtes. On s'aperçoit que le comportement général de cette évolution peut varier de façon extrême pour des valeurs de p proches, par exemple, si l'on essaie de modéliser les propriétés d'un solide dont les liaisons entre les molécules sont détruites au hasard.

Il m'a été difficile de réaliser des expériences sur le modèle B, étant donné que je n'ai pas trouvé de construction associée à la modélisation. En effet, étant donnée une distribution de degrés, rien ne dit qu'il est possible de construire un graphe correspondant (cas évident où la somme des degrés est impaire...), ni combien de graphes sont possibles. Ce défaut est négligeable dans le sens où ce qui intéresse les auteurs de [?] est le rapport entre les prévisions du modèle et les données expérimentales à leur disposition, sans qu'ils cherchent par exemple à réaliser des analyses d'algorithmes. Les informations de [?] et de l'ouvrage de référence [?] sont à ce titre présentées de façon plus *informaticienne*.

Annexes

Formulaire de probabilités

Espace de probabilités On appelle espace de probabilités S un ensemble dénombrable, tel qu'à chaque élément x de S on associe un poids w_x , et tel que $\sum_{x \in S} w_x = 1$. On appelle événement un sous ensemble E de S . On appelle probabilité d'un événement $E : P[E] = \sum_{x \in E} w_x$.

Variable aléatoire On appelle variable aléatoire une fonction X qui associe un réel à tout élément de S . On note $X = k$ pour $e \in S | X(e) = k$. On appelle espérance d'une variable aléatoire X la quantité $E(X) = \sum_k k * P[X = k]$.

Lemme 5.1 (Inégalité de Markov) Soit X une variable aléatoire prenant uniquement des valeurs non nulles. On a :

$$P(X \geq t) \leq E(X)/t$$

De plus, si X ne prends que des valeurs entières, on a : Si $E(X) \mapsto 0$, alors $P(X = 0) \mapsto 1$

Preuve $E(X) = \sum_{k \geq 0} k * p_k \geq \sum_{k \geq t} k * p_k \geq t * (\sum_{k \geq t} p_k) = t * P(X \geq t)$

Fin de preuve

Variance On appelle variance⁷ la quantité : $Var(X) = E((X - E(X))^2)$

Le formalisme suivant permet d'exprimer de façon élégante certains résultats sur le modèle B.

Fonction génératrice d'une variable aléatoire Soit X une variable aléatoire à valeurs entières, non négatives. La fonction génératrice de X est définie par :

$$G(x) = \sum_{n=0}^{\infty} P(X = n) * x^n$$

On supposera par la suite que cette série converge absolument pour tout $x \leq 1$.

Lemme 5.2 (Propriétés des fonctions génératrices) Quelques propriétés usuelles :

- $P(X = k) = \frac{G^{(k)}(0)}{k!}$
- $E(X) = G'(1)$
- $Var(X) = G''(1) + G'(1) - [G'(1)]^2$

Définitions

Diamètre Le diamètre d'un graphe $G = (V, E)$ est défini par :

$$\max_{u, v \in V} (d(u, v))$$

Propriété monotone Une propriété est qualifié de monotone si elle est préservée par addition d'arêtes.

Coefficient de clusterisation Le coefficient de clusterisation est défini comme la probabilité que, partant d'un sommet d'un graphe, et choisissant deux voisins, il existe une arête entre eux.

Avec $\lambda_G(v)$ le nombre de triangles pour dans lesquels le sommet v apparait, et $T_G(v)$ le nombre de triplets que l'on peut former avec deux voisins de v , on a :

$$C = \frac{1}{n} * \sum_{v \in V} \frac{\lambda_G(v)}{T_G(v)}$$

⁷La variance représente la dispersion d'une variable aléatoire autour de sa valeur moyenne

Références

- [1] Reka Albert and Albert-Laszlo Barabasi. Statistical mechanics of complex networks. *Reviews of Modern Physics*, 74 :47, 2002.
- [2] B. Bollobas. *Random Graphs*. Cambridge University Press (2001), 2001.
- [3] M. E. J. Newman, S. H. Strogatz, and D. J. Watts. Random graphs with arbitrary degree distributions and their applications. *Physical Review E*, 64 :026118, 2001.
- [4] D. B. West. *Introduction to Graph Theory (2nd Edition)*. (Prentice Hall, Upper Saddle River), 2001.